3D Face Recognition Using Multicomponent Feature Extraction from the Nasal Region and its Environs

Gao, Jiangning

Award date:
2016

Awarding institution:
University of Bath

3D Face Recognition Using Multicomponent Feature Extraction from the Nasal Region and its Environs

by

Jiangning Gao

The thesis submitted for the degree of

Doctor of Philosophy

in

The Department of
Electronic and Electrical Engineering
University of Bath

October 2016

Signature: . . . . . Date: . . . . . .

Acknowledgements

Firstly, I would like to offer my sincere gratitude to my dear supervisor, Dr Adrian Evans, for his invaluable guidance, consistent support, encouragement and willingness in helping me develop as an independent researcher, including scholarship application, brilliant comments and proofreading my publications. I am sincerely grateful to be one member of his research team who supports me to take part in various research activities in terms of the conference, workshop, seminar, symposium and summer school, which provides me the chance to present my work and achieve more precious suggestions.

I would also like to thank Dr. Nathan Smith, Dr Mehryar Emambakhsh and Dr Mark Hansen for their generous advices, support and helpful discussion.

I would like to thank the National Institute of Standards and Technology (NIST) and University of Notre Dame for providing the access to the Face Recognition Grand Challenge (FRGC) database, Bogazici University for the Bosphorus database, the University of West England for the Photoface database and the 3DE-VISIR database, and Vitomir Struc for his publicly available toolbox for face recognition.

Great acknowledgement is given to the University of Bath Graduate School and China Scholarship Council (CSC) for sponsoring my PhD study. Also, FG'15 DC scholarship and Santander fund for travel bursaries.

I would like to thank my best friends I met in Bath, Miss Yuanyang Rong, Dr. Maomao Zhang, Miss Lian Fan, Mr Zhongjian Liu, Dr. Meng Li, Dr. Lin Zhou, Miss Wentao Deng, Mr Yangzi Liu, Miss Jing Hu, Miss Xi Duan, Dr. Lu Ma, for all the unforgettable moments and emotional support during my PhD study.

Last, but certainly not least, I would like to express my ultimate gratitude to my family. A special thanks to my parents and parents-in-law, for all of their kind support. I would like to express my warm thanks to my beloved husband, Mr Xinhai Zhang, who gives me endless support and encouragement, especially when I felt frustrated and confused in the past four years.

Abstract

This thesis is dedicated to extracting expression robust features for 3D face recognition. The use of 3D imaging enables the extraction of discriminative features that can significantly improve the recognition performance due to the availability of facial surface information such as depth, surface normals and curvature. Expression robust analysis using information from both depth and surface normals is investigated by dividing the main facial region into patches of different scales. The nasal region and adjoining parts of the cheeks are utilized as they are more consistent over different expressions and are hard to deliberately occlude. In addition, in comparison with other parts of the face, these regions have a high potential to produce discriminative features for recognition and overcome pose variations.

An overview and classification methodology of the widely used 3D face databases are first introduced to provide an appropriate reference for 3D face database selection. Using the FRGC and Bosphorus databases, a low complexity pattern rejector for expression robust 3D face recognition is proposed by matching curves on the nasal and its environs, which results in a low-dimension feature set of only 60 points. To extract discriminative features more locally, a novel multi-scale and multi-component local shape descriptor is further proposed, which achieves more competitive performances under the identification and verification scenarios.

In contrast with many of the existing work on 3D face recognition that consider captures obtained with laser scanners or structured light, this thesis also investigates applications to reconstructed 3D captures from lower cost photometric stereo imaging systems that have applications in real-world situations. To this end, the performance of the expression robust face recognition algorithms developed for captures from laser scanners are further evaluated on the Photoface database, which contains naturalistic expression variations.

To improve the recognition performance of all types of 3D captures, a universal landmarking algorithm is proposed that makes uses of different components of the surface normals. Using facial profile signatures and thresholded surface normal maps, facial roll and yaw rotations are calibrated and five main landmarks are robustly detected on the well-aligned 3D nasal region. The landmarking results show that the detected landmarks demonstrate high within-class consistency and can achieve good recognition performances under different expressions. This is also the first landmarking work specifically developed for the reconstructed 3D captures from photometric stereo imaging systems.

Content

List of Figures:

List of Tables:

Acronyms

3DE-VISIR	3D Expression-VISible and near-InfraRed database
AAM	Active Appearance Model
AFM	Annotated Face Model
BSFS	Backward sequential feature selection
CMC	Cumulative Match Characteristic
DCT	Discrete Cosine Transform
EER	Equal Error Rate
EGI	Extended Gaussian Images
FACS	Facial Action Coding System
FAR	False Acceptance Rate
FRGC	Face Recognition Grand Challenge
FRR	False Rejection Rate
FSFS	Forward sequential feature selection
g	Gallery
GA	Genetic algorithms
H	mean curvature
HMM	Hidden Markov models
K	Gaussian curvature
KFA	Kernel Fisher's Analysis
k_{max} **and** k_{min}	Principal Curvatures

ICA	Independent Components Analysis
ICP	Iterative Closest Point
LBP	Local Binary Patterns
LDA	Linear Discriminants Analysis
LOO	Leave-one-out
IR	Near infrared light
NMF	Nonnegative Matrix Factorization
NN	Nearest Neighbours
p	Probe
PCA	Principal Components Analysis
PS	Photometric stereo
R_1RR	Rank one Recognition Rate
ROC	Receiver Operating Characteristic
RR	Recognition Rate
SFR	Spherical Face Representation
SI	Shape Index
SIFT	Scale Invariant Feature Transformation
SNs	Surface normals
SVM	Support Vector Machine
VIS	Visible light

Chapter 1 Introduction

1.1 Background and Challenges

Biometrics that try to utilise some physical and behavioural characteristics to represent the human beings have become more significant in the recent decades, and have been widely used in security, control and entertainment applications. With the increasing security concerns of access control, for example of military, airport, bank and border control, the demand to develop robust biometrics features for recognising and verifying people has been the subject of much research interest.

In recent years, some smart devices have been developed for many purposes and scenarios, which can produce many diverse biometric applications. For example, fingerprints have been widely used to facilitate the access control of most smart phones and, to achieve a high security, the use of the iris has been proposed [1]. Furthermore, the face has been introduced to bank payment system, for example by the HSBC bank, which is aimed to replace the time-consuming or even less accurate manual recognition by human. Therefore, to accommodate the increasing security demands, more sophisticated and robust biometrics are required to improve the performance.

Compared to behavioural biometrics, physical biometrics (e.g. fingerprints, iris, pupil and face) are assumed to be unique for each identity and can achieve high recognition performance. In particular, fingerprints and iris have been widely used in many scenarios as their high accuracy and distinctness. However, the main problem of these two modalities is the requirement of human cooperation during data acquisition, which might not be appropriate in unconstrained scenarios. In contrast, facial information acquisition is more convenient as the development of low-cost capture devices and the main parts of face are hard to be intentionally or unintentionally occluded by other things. Also, the human face maintains a lot of discriminative features for recognition, which brings more possibilities to use the face as a biometric for security concerns and the development of diverse software applications.

Face recognition techniques mainly come from computer vision, pattern recognition and artificial intelligence and employing the face as a biometric feature has been used in many application scenarios [2]. In its early days, most face recognition algorithms used 2D intensity

or colour images. As the accuracy of 2D face recognition is adversely affected by variations in pose, expression, illumination, and occlusions, the development of robust, fully-automatic 2D face recognition system remains a challenging problem. Alternatively, using 3D facial data to understand the characteristics of the human face in the 3D domain has been shown to have much promise for improving the overall recognition performance.

Using 3D face as a biometric has many advantages in comparison with 2D imaging: (1) 3D face shape possesses high distinctiveness among different identities, which can be extracted from various types of 3D shape representations in terms of depth, surface normals, curvatures and so forth. Some specific regions, for example the nose, have been shown to contain lots of discriminative features [3-5]. (2) 3D acquisition systems have the potential to solve the illumination problem of 2D imaging as feature detection and the consistency of feature extraction by 2D captures fail in different illumination conditions. (3) Compared to 2D methods, 3D facial shape provides more possibilities to address the problems caused by expression, pose and occlusion variations. Feature extraction on some relatively rigid regions, for example of 3D nose or certain small facial patches, is one of the most popular approaches to address expression variations [6, 7]. Also, it is easy to calibrate pose variations on the 3D shape and the missing areas caused by occlusions can be estimated from the known areas.

Although 3D strategies have the potential to address many drawbacks of 2D imaging and improve the recognition performance of biometric systems, there are still many challenging and unsolved issues in 3D face recognition, including the expression robust feature extraction, low complexity and accurate pose calibration, and real world applications. Expression invariant 3D face recognition requires developing more effective and efficient descriptors for feature extraction on the 3D shape or evaluating some rigid patches on the face, which suffer less facial surface changes or movements under expressions. For the pose calibration, extracting pose invariant features and making a compromise between the efficiency and accuracy are two hot topics in 3D face recognition.

To overcome those challenges in a more effective and efficient way, many researchers have proposed to investigating the 3D nose for recognition as it demonstrates high robustness under various expression variations. Also, it is the most symmetric and obtainable region on the face, which provides more facial surface information for pose calibration and discriminative features for recognition. Therefore, there is increasing research of interest on utilising the robustness and distinctness of nasal region for alignment and feature space manipulation.

Most public 3D face recognition algorithms are tested on the depth map with higher accuracy, which are mainly reconstructed from 3D point clouds captured by laser scanners or structured light. The main drawback of these kinds of acquisition methods is that they require candidates' fully cooperation as well as long calculation time to produce the 3D captures, which might be hard to be applied to some security control scenarios. Therefore, the development of more practical 3D acquisition system and discriminative features extraction from the unconstrained and low-complexity 3D captures have become a hot area in face recognition.

1.2 Motivation

1.2.1 Expression Robust Feature Extraction on the Nasal and Adjoining Cheek Regions

Compared to other biometrics modalities, face recognition possesses many distinct advantages and 3D imaging provides more potential to address the problems caused by facial variations and illumination conditions, which are hard to solve by using the 2D imaging. However, variations in expression, which lead to muscle movements and deform the face surface, still present challenges to recognition systems using 3D data. Therefore, one of the main working areas in this thesis is extracting relatively expression robust features for 3D face recognition.

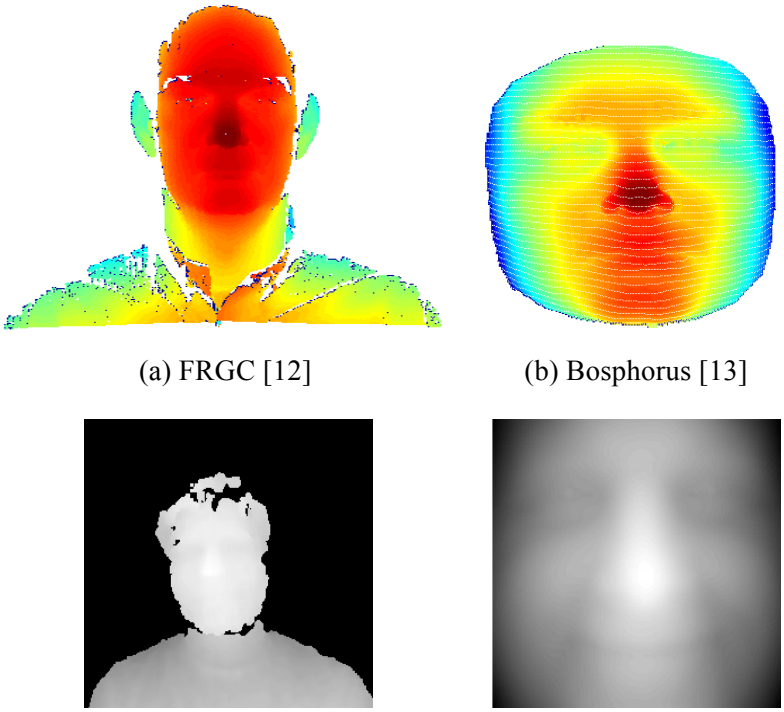
One way to address these problems is to select relatively stable structures or patches on the facial surface and use these to explore the discriminative features. For example, the nasal region, which is more consistent over universal expressions and also invariant to the majority of occlusions caused by hair, hands and scarves, is widely used in many face recognition algorithms. In addition to the nasal region, the adjoining parts between the nasal region and cheek bones also have the potential to produce discriminative features [8], even though they are less expression robust than the main nasal region. As a consequence, it is interesting to extend the nasal region to its surroundings to extract more discriminative patches or design new descriptors for investigating on the nasal and adjoining cheek regions.

1.2.2 Investigation of Photometric Stereo Captures

Another main motivation of this topic is improving the real world application possibility, for example of utilising 3D captures from the photometric stereo. In general, to achieve high identification and verification performance, many researchers take efforts to investigate the use of the depth maps with higher accuracy in facial details, for example the captures provided by the FRGC and Bosphorus databases in Figure 1.1. Many algorithms can produce high rank one recognition rate (R_1RR), even higher than 99%. However, those methods lose the real world

application feasibility as nearly all the captures they used are obtained from laser scanner or structured light, which are not suitable to apply to most scenarios for security concern, even though the resulting 3D captures possess higher resolution and accuracy and provide a good source for face analysis.

In the past decade, many new 3D data acquisition devices have been developed. The most famous one is the Kinect device developed by Microsoft, which has been widely used in Xbox. Since the depth information is relatively with low resolution as shown in Figure 1.1(c), many researchers proposed to merge both 2D colour and depth information, defined as 'RGB-D', to design face recognition algorithms. The Photoface device [9] proposed by Centre of Machine Vision at University of the West of England provides a good compromise of this tough and challenging issue. The Photoface device facilitates the 3D data acquisition procedure, which can obtain the facial data in a very short period of time without candidates' cooperation. What is more, the resulting 3D captures provide a good overall reconstruction in comparison with the "ground-truthed" captures, which produces relatively higher resolution 3D captures than other commercial acquisition system [10]. As shown in Figure 1.1(c), the depth information of Photoface captures used in this thesis is recovered by integrating the surface normals using the Frankot-Chellappa surface integration algorithm [11].



(a) FRGC [12] (b) Bosphorus [13]

(c) EURECOM Kinect [14] (d) Photoface [9]

Figure 1.1: Example 3D captures laser scanner (FRGC), structured light (Bosphorus), Kinect sensor (KinectFaceDB) and Photoface.

Actually, there has been limited research using the Photoface captures for face recognition, although this kind of data possesses a lot of advantages and has promising application prospects. It is interesting to apply the well-designed face recognition algorithms to the Photoface captures or investigate new techniques on the basis of their characteristics. Therefore, this work will consider the development of new feature detection and extraction approaches to achieve high recognition performance under real world application scenarios.

1.2.3 Landmarking for the Captures with High/Low Accuracy Depth Maps

In biometric systems, landmarks are often instrumental in the generation of signatures for faces [15]. Landmarking plays a significant role in region based face recognition algorithms and, as such, its accuracy will directly determine the effectiveness of feature extraction. The within-class similarity and between-class dissimilarity may be adversely affected by inconsistent landmarks. In addition, for real-time automatic face recognition systems, the landmarking method should also be computationally efficient. Landmarking has been well researched and explored using higher accuracy depth maps as many facial details are preserved, for example see [16]. However, the landmarking approaches designed for the captures produced by newly developed and more practical 3D data acquisition systems have not been explored in literature, which is a major impediment to the development of region based face recognition algorithms.

3D data acquisition in a biometric session is the first main step of a 3D face recognition algorithm and the use of different acquisition devices can result in captures with varying characteristics, presenting different challenges for landmarking. Many landmarking approaches use databases captured using laser scanners or structured light technologies, which provide point clouds captures with high accuracy of depth maps and better preserve the face's local details. However, such acquisition systems require full cooperation during data acquisition, which presents difficulties for real world application scenarios. The Photoface device [9, 17] that collects 3D captures by the photometric stereo was proposed to address this problem. In addition to the real time and low complexity data acquisition, the resulting captures are well smoothed with high resolution, which have been proven to be more beneficial to face recognition in comparison with other kinds of captures, as more facial fine discriminative details are reconstructed on the depth maps [18]. However, the depth maps produced from surface normals are flatter than their real-world counterparts with most protrusions understated [18], such as the nose tip, eye sockets and the sides of the nose.

Those regions with the highest curvatures are crucially important for nasal landmarks localization. Therefore, large variations on those regions might result in failed landmarking on the Photoface database using the curvatures based algorithms. Region based face recognition algorithms are still unexplored on this kind of data due to the lack of effective landmarking algorithms, which is also the gap in the 3D landmark detection literature. Simply applying well-designed landmarking algorithms to this kind of data is not guaranteed to produce a good performance. Therefore, new landmarking algorithms designed for such kind of 3D captures should be further explored.

1.2.4 The Use of Surface Normals

Most 3D face recognition algorithms are focused on utilising the depth information to detect and extract features. In addition to the depth, the surface normals of each point can determine the orientation of facial surface and describe the local shape variations, which provides an effective facial surface representation to investigate discriminative features for recognition. Gökberk *et al*. [19] found that surface normals provide the best features for face recognition in comparison with the depth, point clouds and shape index.

Surface normals (SNs), that can be represented as a vector of three individual components (SN_x, SN_y, and SN_z), have been widely used in feature detection and extraction [6, 9] on the 3D captures with higher accuracy, providing the motivation for fusing the depth and surface normals to exploit more discriminative features for 3D face recognition. For example, Li *et al*. use the 3D point clouds to estimate the surface normals and extract different scales of local normal patterns on three components of surface normals [6]. In [20], the surface normal vectors at the sampled facial points are shown to contain more discriminatory information than the coordinates of the points.

In addition, using surface normals is beneficial to feature space manipulation on the 3D captures from the photometric stereo as the original data computed from the photometric stereo are surface normals and the albedo, which are less smoothed and suffer less distortion than the reconstructed depth information. Zafeiriou *et al*. [9] utilize the surface normals as vectors and match the normalface for recognition. The recognition performance achieved by normalface outperforms that using depth information. Therefore, using surface normals as an effective 3D shape representation has a lot of advantages for feature space manipulation, providing an additional feature source for expression robust face recognition [6]. In particular, the surface

normal has great potential to provide a good approach for the landmarking on the Photoface captures.

1.3 The Outline of This Thesis

In Chapter 2, an overview of the most significant research work related to 3D face recognition algorithms in the literature is provided. An introduction to biometrics and face recognition is first given in Section 2.1. Then the basic 3D face recognition techniques are introduced in Section 2.2, including preprocessing, landmarking, features extraction, postprocessing, matching and decision making. In particular, the recognition algorithm for addressing the expression variations is further explained in Section 2.3 and feature extraction on the nasal and adjoining cheek regions is introduced and discussed in Section 2.4. In addition, to demonstrate the advantages of using 3D nasal and adjoining cheek regions, three whole face evaluations on the Bosphorus database are provided in Section 2.4.4. A summary and the challenges are concluded in Section 2.5.

In Chapter 3, the use of 3D face databases is further explained, including an overview of 3D face database, preprocessing work and baseline performance evaluation. In Section 3.2, an analysis of all widely used 3D face databases is introduced, which summarises all the significant characteristics and challenges and provides a useful reference for 3D facial analysis. Then, the motivation and experimental settings of the databases (FRGC, Bosphorus and Photoface) used in the following chapters are further explained in Section 3.3. The preprocessing work and the recognition performance evaluations on a subset of three databases are provided in Section 3.6.

In Chapter 4, an expression robust rejector is proposed that first robustly locates landmarks on the relatively stable structure of the nose and its environs, termed the cheek/nose region. Then, 75 nasal curves and 38 cheek/nose curves are found by concatenating sets of landmarks. For manipulating the feature space, 113 curves are selected and then, in Section 4.6, after some significant parameters analysis, a small set of features (4 curves with only 15 points each) on the cheek/nose surface are finally selected to build the feature space. In Section 4.7, the resulting pattern rejector, which can quickly eliminate a large number of candidates at an early stage, is further evaluated on the FRGC database for both the identification and verification scenarios. The classification performance using only 60 points from 4 curves shows the

effectiveness of this efficient expression robust rejector. In addition to the horizontal curves, the radial and root curves are also explored in Section 4.8.

In Chapter 5, a novel local depth and surface normals descriptor is proposed. This is used to explore the discriminative features on the nasal surface and adjoining cheek regions for expression robust 3D face recognition. After preprocessing the 3D face data, landmarks located on the perimeter of a triangular region covering the nose and adjoining parts of the cheeks are accurately detected. Inspired by Local Binary Patterns, local shape differences for 3D points on a set of horizontal curves joining selected landmarks provide a novel representation of the local shape information. A further analysis of the discriminatory power of each patch shows that the adjoining regions have the potential to produce good recognition performance. Using the FRGC and Bosphorus databases, the performance of the proposed descriptor is evaluated on diverse patches, scales and for four components, one from the depth and three from the surface normals. Results show that the new local shape descriptor performs well at representing the shape information on a relatively large scale. On the basis of this descriptor, a relatively small set of features is extracted from the nasal and adjoining cheek regions. The adjoining cheek regions demonstrate a high discriminatory power and provide a useful new addition to 3D face biometrics.

In Chapter 6, an investigation of expression robust features extracted from the Photoface captures is provided to demonstrate the recognition performance of reconstructed data using improved well designed algorithms for fine 3D data. 75 curves are found by connecting 16 automatically detected landmarks on the nasal region which result in an efficient feature set consisting of different 3D data representations, depth, surface normals and shape index. The recognition performance is tested on the Photoface, FRGC and Bosphorus databases and features extracted using surface normal y component (SN_y) and shape index map from the Photoface database significantly outperform the other databases. After feature selection, a small number of features are tested on the 3DE-VISIR Database using one training sample scenario. The recognition performance shows that nasal curves extracted from SN_y and shape index map are relatively expression robust which can be used for 3D nose recognition.

In Chapter 7, a new 3D facial landmarking algorithm is proposed that is applicable for both higher accuracy 3D captures, obtained using laser scanners or structured light, and the reconstructed captures from photometric stereo systems. The benefits of employing surface normals are demonstrated for both roll and yaw rotation calibration and nasal landmarks

detection. Results show that the surface normals can provide an additional and effective representation of the 3D surface geometrical structure. By thresholding the surface normal maps, the nasal bridge and candidate alar grooves are robustly detected, which can be utilized to refine the location of nose tip and correct rotation errors in the yaw direction. Also, using the nasal bridge, root and nose tip detected by the surface normal maps, a coarse to fine roll rotation calibration is applied to further refine the pose variations, even though using surface normal maps is relatively pose invariant. Finally, five landmarks on the nasal region are robustly detected on the well aligned 3D captures by utilising the thresholded surface normal maps and facial profile signatures. This novel landmarking algorithm provides an effective way to localize the key landmarks on the human nose. Test results on the Bosphorus and FRGC v.2 databases, which contain a large number of expression variations, show that the detected landmarks possess high within-class consistency and produce good recognition performance under different expressions. This landmarking algorithm is further applied to less accurate 3D facial captures reconstructed from infrared photometric stereo captures and is shown to robustly detect the main nasal landmarks are robustly detected. The proposed approach is the first landmarking work specially designed for this kind of 3D captures.

The thesis concludes with chapter 8, which provides a summary of the proposed algorithms in section 8.1 and discussion and future work in section 8.2.

1.4 Main Contributions

The main contributions of the thesis and its related field of science can be categorised as follows:

1. An overview of public 3D face database in terms of database population, data acquisition, preprocessing and landmarking and modes of facial variations is introduced for the first time. It provides an appropriate reference for the choice of 3D face databases in 3D facial analysis.

2. Using both depth and surface normal information, a recognition performance analysis of the main parts of human face is explained for the first time, including within-class dissimilarity under expressions, large scale and local patches evaluation.

3. A low complexity pattern rejector for expression robust 3D face recognition by matching curves on the nasal and adjoining cheek region is proposed, which provides the motivation of investigating the discriminative features on the adjoining cheeks.

4. A novel local shape descriptor is proposed to explore the discriminative features on the nasal surface and adjoining cheek regions, which outperforms the nasal curves. Different 3D shape representations in terms of depth and three components of surface normals are exploited.

5. An investigation of expression robust features extracted from the Photoface captures is provided to demonstrate the recognition performance using the well-designed algorithms for 3D captures with higher accuracy depth maps. A comparison of recognition performances among the FRGC, Bosphorus and Photoface databases shows that the Photoface captures have the potential to extracting discriminative features for recognition but use the different feature space manipulation methods.

6. Utilizing the surface normals maps and facial profile signatures, a novel 3D facial landmarking strategy is proposed for the 3D captures with higher or lower accuracy depth maps. Roll and yaw rotation calibration achieved by thresholding the surface normal maps and the detected nasal bridge and landmarks.

Chapter 2 Literature Review

2.1 Biometrics and Face Recognition

2.1.1 Biometrics

Biometrics is the use of science and technology to obtain behavioural or physical characteristics of a person for identification or verification purposes. The main motivation of storing a person's biometric data is to reuse it for the recognition purposes. Due to the wide variety of biometric modalities, there are many data acquisition methods and analysis algorithms. In general, the biometrics of human beings are derived from physiological and behavioural features [21]. Fingerprints, irises and pupils, faces, palm print, voice, keystroke and gait are among the most widely used human biometrics.

Gait, voice and keystroke are typical behavioural features of biometrics. The main disadvantage of this kind of method is the requirement of high level feature detection algorithms. In addition, behavioural features are hard to obtain accurately and straightforwardly in real world applications and can also change over time. Therefore, the use of behavioural features as biometrics may not be reliable enough for high security applications. However, behavioural characteristics have found applications in soft biometrics.

Physiological biometrics technologies are widely used in both academia and industry as this kind of features is more discriminative and possess high within-class similarity and between-class dissimilarity. The face, fingerprints and iris are the most widely applied biometrics due to their high distinctness. Iris patterns have been proven to be unique for each person and, as a biometric, possess a very low false acceptance rate (FAR) [22]. However, simply comparing the FAR with other recognition approaches such as the face or fingerprints does not show the full picture. For example, iris recognition has a high false rejection rate (FRR) resulting from eye blinking and the use of contact lenses. Although the iris requires relatively high cost devices for real world applications, its high accuracy has promising application prospect in the scenarios with high security requirements.

In comparison, fingerprints have been widely used in access control systems such as building entrances, border controls, most smartphones and laptops due to its unparalleled efficiency and effectiveness. The main drawback of fingerprints is that they generally require candidates to

cooperate by making physical contact with the sensor surface. As a biometric modality, the human face is advantageous as it is an easily collectible and nonintrusive physiological biometric, compared with the iris and fingerprints [23]. The face has contactless acquisitions and can be widely applied to the surveillance systems [2].

Jain *et al*. define some good characteristics of biometric systems [24, 25]: for the data acquisition system, the first issue is convenience and efficiency, which require biometrics features in both probes and gallery should be easy and fast to obtain. Some complicated data acquisition devices might degrade the practicability of a biometric system. Also, safety is another key issue for data acquisition. The data acquisition process is completely private for a candidate subject and the biometric data should be securely stored or encrypted to avoid any spoofing attack. The extracted biometric features, either physical or behavioural, should be discriminative, consistent and robust, which can best preserve the within-class similarity and between-class dissimilarity. They require that the features are hard to be falsified by the subject and also with high consistency under various scenarios.

In general, there are two exclusive biometric application scenarios, verification and identification, as shown in Figure 2.1. For the verification or authentication scenario, a one-to-one matching of the probe data and the gallery data of the same identity is undertaken, in which the subject's identity is compared with the claimed identity. A threshold is chosen to compare with the matching score and determine whether the target subject is accepted [25]. The verification performance is measured by the Receiver Operating Characteristic (ROC) curve which plots the FRR versus the FAR or the verification rate (1-FRR) versus the FAR [26]. The FAR is defined as the percentage of the probes that are wrongly recognized as the claimed person while the FRR is the percentage of the probes that are incorrectly rejected. Therefore, the ROC curve describes the trade-off between the FAR and the FRR. EER is the point where the FRR equals the FAR and it is the most commonly stated single number on the ROC curve. In addition, the FRR or verification rate at 0.1% FAR on the ROC is also used to evaluate the performance.

For the identification scenario, the input subject is compared with the whole database and the closest subject is found to denote its identity. The Cumulative Match Characteristic (CMC) curve is applied to evaluate the identification performance and plots the recognition rate (RR) versus the rank number, which summarizes the percentage of probes and galleries that are correctly matched [26]. R_1RR is the most widely used performance metric on the CMC curve.

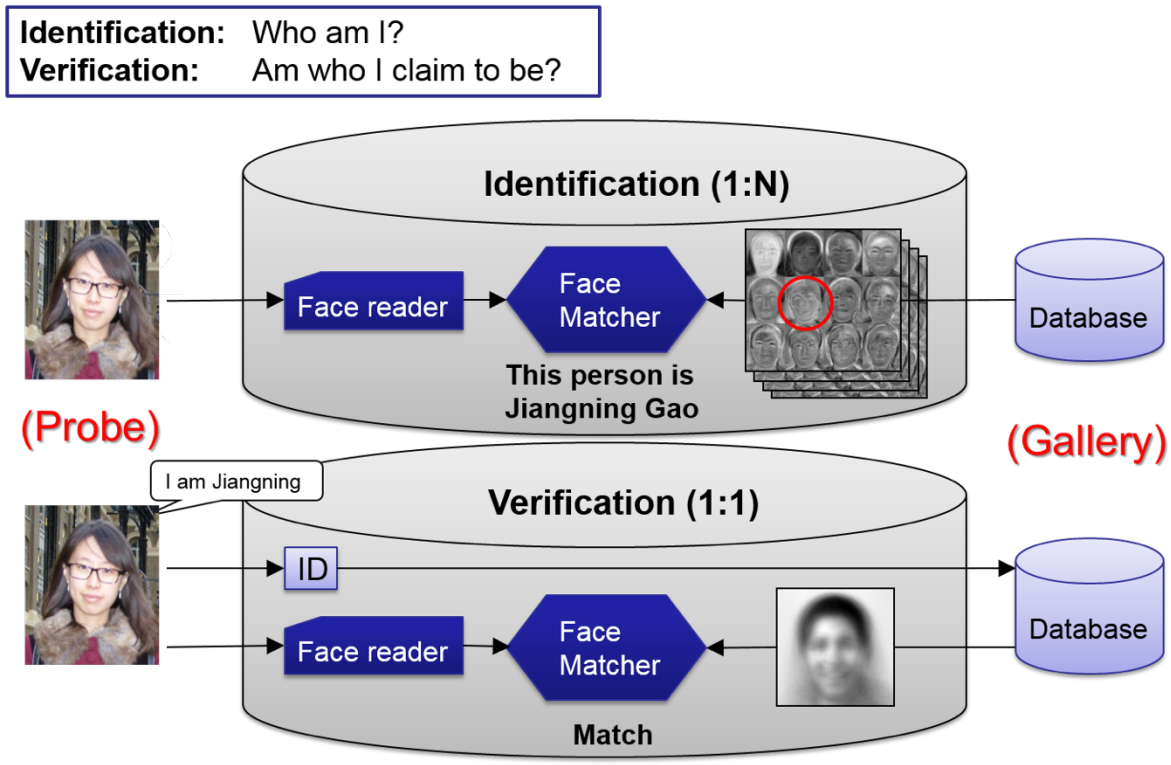


Figure 2.1: A description and comparison of face verification and identification system. For identification, the probe is identified by matching all the biometric data in the gallery and the output is the closest identity. For verification, the claimed identity of probe is verified by matching with its corresponding biometric data in the gallery.

In the all-to-all matching scenario, another method used for identification is 'leave-one-out' cross validation, which considers each capture as an individual and compares the target capture with the remaining captures by calculating a distance measure. The nearest subject denotes the probe's identity and the Nearest Neighbours (NN) are widely used to produce the matching score or distance.

2.1.2 Face Recognition

As a branch of object recognition, face recognition has been widely investigated over several decades. The main motivation of using face as a biometric is the diversity of human face, which can provide significant contributions in improving the performance and applicability of biometric systems. Facial captures contain plenty of discriminative information, which is relatively robust for each subject. For example, people may possess various types of eye, nose, eyebrow, forehead, cheek and mouth, which bring many different shape features of each facial part or their combinations. Considering behavioural features, some expressions (e.g. smile,

disgust or frown) on the human face also contain discriminatory features for each person. All those diversities result in the special distinctiveness of the human face and many holistic or region based recognition algorithms are proposed to extract the facial features.

Confronted by the increasing security concerns, the use of biometric features for recognising people in many application scenarios has become more important. Although some existing biometric technologies (e.g. fingerprints and irises) can offer a very high accuracy, they also require much greater explicit cooperation from the user. As one of the most commonly used biometrics, the face can be easily acquired by imaging devices in realistic scenarios. Thus there is a significant application-driven demand for improved performance in face recognition [27].

Using the face as a biometric has significant application prospects. First, the data acquisition procedure is less intrusive and more straightforward in comparison with the iris and fingerprints. It is hard to intentionally or unintentionally occlude the whole face during the acquisition. Also, imaging modalities are generally much cheaper than other biometrics and can produce facial captures with good quality [27]. In addition, the computation cost of image processing algorithms has also been reduced. Instead of some traditional biometric application scenarios, face recognition is exclusively used in identity card organization systems in many countries. The photo capture of each citizen is always the first collected biometric, and robust face recognition algorithms also contribute the personnel screening and criminal suspect decision.

2.1.3 The Motivation of Using 3D Face Recognition

With the developments in computer vision and pattern recognition fields, many algorithms have been applied to face recognition systems. 2D images are widely used in face recognition as it is easy to obtain discriminative facial features from the texture of 2D grey or colour face captures. However, as can be seen from Figure 2.2, 2D approaches contain some drawbacks in terms of the sensitivity to variations in the illumination, pose, facial expression and occlusion, which can dramatically affect the within-class similarity and between-class dissimilarity in each 2D capture.

In the past two decades, many the state-of-the-art 3D face recognition techniques have been proposed to solve the aforementioned issues found in the 2D imaging. The most significant one is the improvements of 3D imaging capture devices in terms of laser sensors, structured light and photometric stereo imaging. These provide well-known approaches to collect the facial surface data and estimate 3D captures. The captures are either stored as mesh grid points or

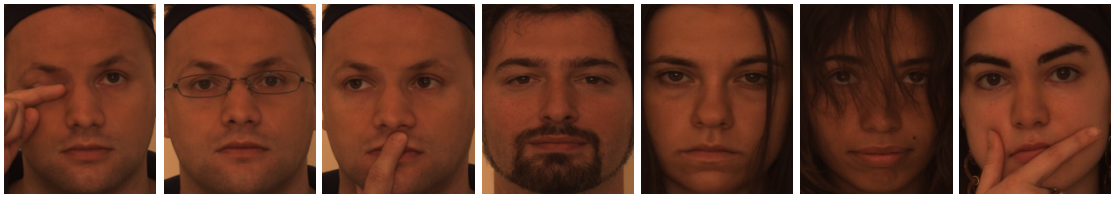
point clouds. The point clouds are widely used in the public databases, for example of the FRGC [12] and Bosphorus databases [13].

3D face recognition has many advantages, which are widely utilised in biometrics. Also, 3D captures have been shown to have the potential for achieving better recognition performance than 2D captures [28-31]. 3D captures can overcome some limitations of 2D face recognition due to viewpoint and lighting variations and also have the advantage of being able to capture shape variations irrespective of illumination variability [29, 31]. In addition, 3D information makes it possible to cope with pose variations by using the curvature information of the face, something not possible with only 2D information [32].

(a) Illumination

(b) Pose

(c) Expression



(d) Occlusion

Figure 2.2: Some facial captures with the illumination, expression, pose and occlusion variations. (a) Four images captured from different light directions by photometric stereo [9] are used to show illumination variations. 21 images with different facial variations in (b) pose, (c) expression and (d) occlusion from the Bosphorus database [13].

The advantages of using 3D face recognition are:

- **Pose calibration and occlusion reconstruction**. In real world applications, the human face might be intentionally or unintentionally rotated by the natural posture and such rotations can dramatically damage the consistency and accuracy of the extracted features. It is relatively easy to correct the facial roll rotation in the 2D images. However, yaw and pitch rotations, which result in missing facial data during data acquisitions, are hard to be calibrated by the 2D imaging. 2D face information is not sufficient to correct the pose around the yaw and pitch directions. 3D approaches can provide more probabilities to address pose variations as they have the potential to approximate the missing points caused by such self-occlusion. Moreover, compared to 2D imaging, some rotation invariant surface descriptors can be found in 3D or 2.5D data.

- **More diverse data representations**. In addition to the texture information used in the 2D imaging, 3D imaging provides the access to the facial surface information such as the depth, surface normals and curvatures. These surface representations provide the possibility to design different descriptors for extracting various discriminative features. By fusing these modalities, the recognition performance can be significantly improved. Also, the curvature calculated from the 3D surface are helpful for face segmentation, feature detection and landmark localization.

- **More robust performance for different expressions**. For both 2D and 3D captures, the variations in expressions are another challenging issue in the recognition. Different degrees and types of expressions might cause the movement of facial patches, which makes the subjects lose the within-class similarity and between-class dissimilarity. Compared to 2D imaging, 3D data has the potential to address this problem as it provides more diversity of data representations. Some descriptors are relatively less sensitive under different expressions. Also, certain regions on the facial surface are comparatively more robust to expression variations, for example of the nose and forehead. Such regions on the 3D map are more consistent under expressions and also contain plenty of discriminative features for recognition in comparison with the 2D imaging.

2.2 Overview of 3D Face Recognition Algorithms

In general, a 3D face recognition algorithm can be divided into four main stages, consisting of the preprocessing, feature space manipulation, matching and classification and decision making, as shown in Figure 2.3. In the preprocessing stage, the main facial part is first cropped from the detected face and then the facial captures are denoised and aligned. The second stage, manipulating the feature space, is crucial in the whole 3D face recognition and includes feature detection, feature extraction and postprocessing. The last two stages are classification and decision making, which have been well explored in many pattern recognition issues.

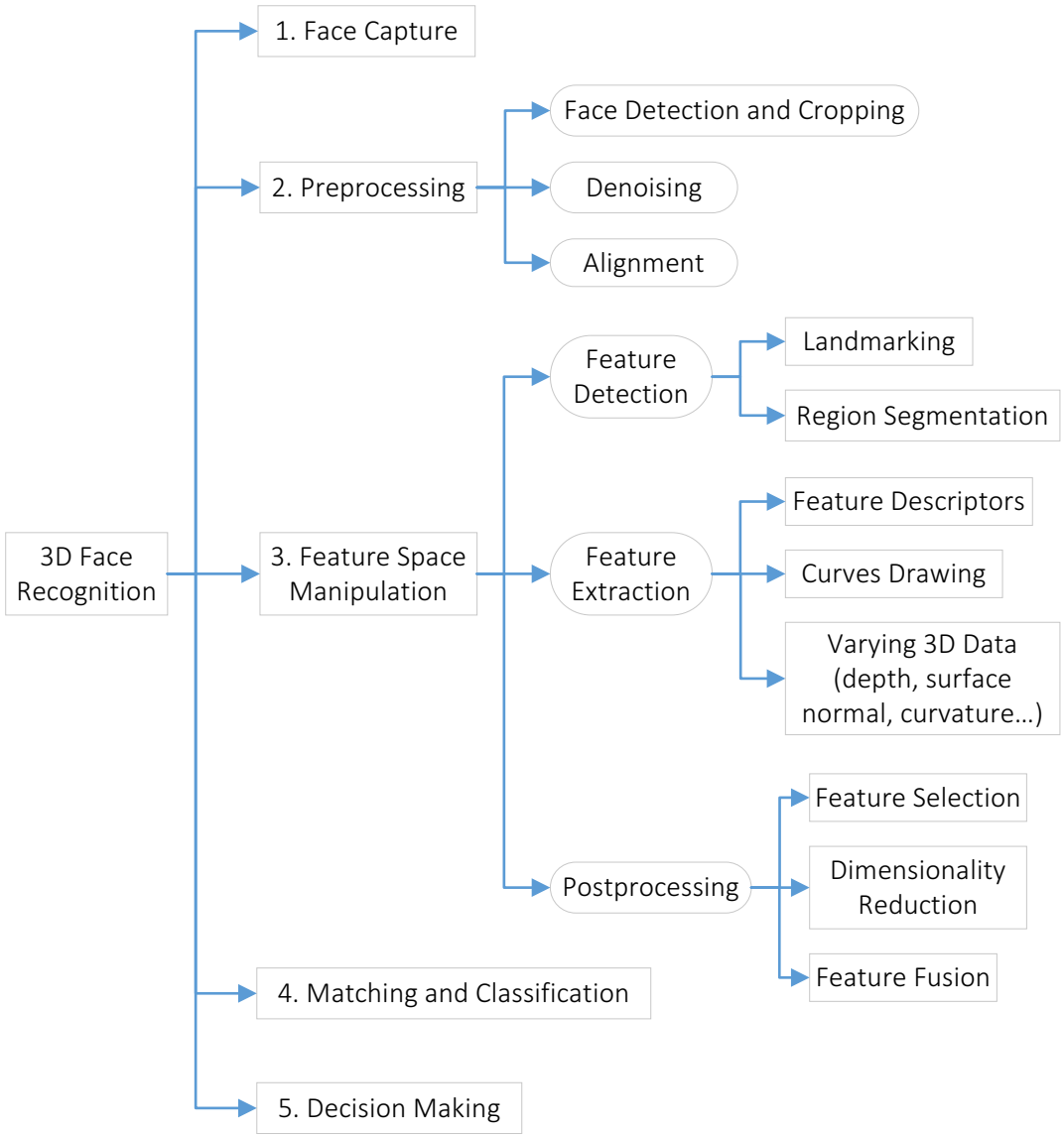


Figure 2.3: The breakdown of a typical 3D face recognition algorithm

2.2.1 Curvatures

3D imaging provides the ability to obtain facial curvature information that facilitates the facial surface analysis. Curvature information plays a significant role in 3D face recognition which has been widely applied in face detection, image segmentation and landmark localization. Also, curvature calculation is roll rotation invariant and not sensitive to the point cloud's storage order.

In differential geometry, two principal curvatures ($k_{\max}$ and $k_{\min}$) that are the maximum and minimum values of the curvatures at a given point of surface measure the amounts of surface bending in different directions, which provide a significant shape operator for facial surface analysis. The Gaussian curvature (K), mean curvature (H), shape index (SI) and curvedness can be estimated using the following equations, which are widely used in the surface analysis.

$$K = k_{max} \times k_{min}$$

$$H = \frac{1}{2}(k_{max} + k_{min}) \tag{1}$$

$$SI = \frac{2}{\pi} \arctan\left(\frac{k_{max} + k_{min}}{k_{max} - k_{min}}\right)$$

$$Curvedness = \sqrt{\frac{k_{max}^2 + k_{min}^2}{2}}$$

Four types of regions are detected by thresholding the curvature maps, which can be employed for region segmentation and feature detection.

1. Convex region: $\begin{Bmatrix} K > 0 \\ H > 0 \end{Bmatrix}$
2. Concave region: $\begin{Bmatrix} K > 0 \\ H < 0 \end{Bmatrix}$
3. Saddle region: $\begin{Bmatrix} K < 0 \\ H > 0 \\ k_{max} + k_{min} > 0 \end{Bmatrix}$
4. Saddle region: $\begin{Bmatrix} K > 0 \\ H > 0 \\ k_{max} + k_{min} < 0 \end{Bmatrix}$

The principle curvatures are very robust for feature space manipulation as they do not change when the surface rotates or translates. For misaligned captures, using other 3D representations might cause greater within-class dissimilarity, but the principle directions are more robust. Also, principle curvature directions have been proven to increase the identification performance significantly as they have high potential to produce discriminative features [33].

The shape index can be computed from the principal curvatures. As a 2D representation of the curvature of 3D surface, shape index is a popular method to characterize the surface patches and many nose, ear and face recognition related algorithms employ it successfully for fiducial feature detection. The shape index describes regions of a surface ranging from spherical cup (-1) to spherical cap (+1) with saddle at 0. SI has been used for landmarking in many algorithms. For example, the nose tip is located in the largest convex region on the face, which can be found by thresholding the SI using $-1 < SI < -5/8$ to produce a binary image [7, 27, 34].

Curvedness is first proposed in [35], which describes the second order structure of the surface in the neighbourhood of any one of its points. Compared to the principle curvatures or Gaussian and mean curvature, curvedness is a positive number that specifies the amount or the intensity of the curvatures. It is defined by calculating the distance from the origin in the principal curvature plane.

2.2.2 Face Detection and Segmentation

Face detection techniques determine the locations and sizes of human faces in digital images or video sequences. In recent years, this form of biological identity verification has been widely applied in the visual surveillance and tracking [36]. For face recognition algorithms, face detection is always considered as the first part of preprocessing procedure because nearly all the captures used in face recognition contain complex backgrounds, or other regions that are not relevant for recognition, for example the neck and hair.

Some efficient and accurate face detection methods are used in recognition algorithms. The most popular and effective methods are based on the colour or curvatures. For colour images, as human skin colour of all races has a specific range in the YCbCr colour space, the skin can be extracted straightforwardly by colour thresholding [4]. Then, the facial region is found by utilising a priori knowledge of facial geometry. For 3D shape, different types of curvatures including the Gaussian curvature, mean curvature and Shape Index are widely used for face

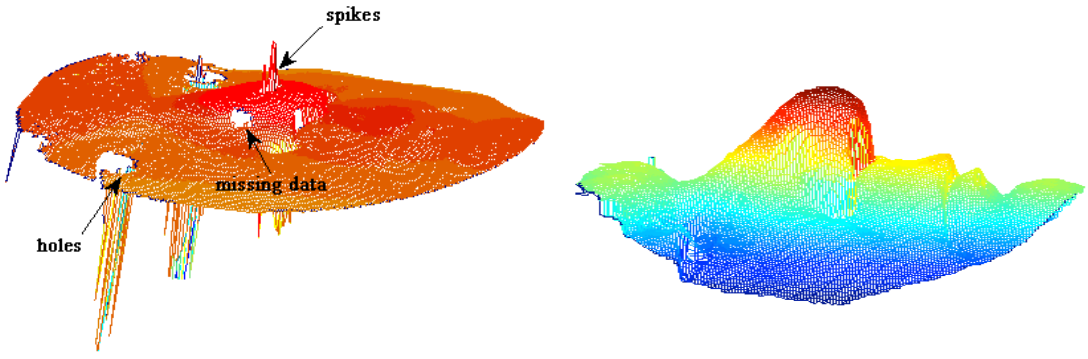
detection [37] due to their rotation invariance [38]. According to a priori knowledge, the facial patches can be found by different thresholds.

In most 3D face recognition algorithms, the nose tip, which is one of the most important landmarks on the face surface, is first detected by calculating the biggest connected convex region. Then, the face region is cropped by a Euclidean sphere centred at the nose tip and the radius of the sphere is chosen by the resolution maps found from the point clouds. The radius is generally set to 75 mm or 80 mm for FRGC [12].

2.2.3 Denoising

Most 3D face captures from various acquisition systems are always inevitably affected by noise to some extent. Noise introduces artefacts to the resulting 3D captures and can cause a significant influence on the overall 3D face recognition performance. As shown in Figure 2.4 (a), spikes are one of the common noise types, which mainly result from impulsive values and can also be interpreted as outliers. Also, the facial surface might contain some holes or valleys. Missing data is another common problem associated with 3D imaging systems, which mainly results from self-occlusion, large depth variations or imaging device inaccuracy.

For most 3D face recognition algorithms, three filters are typically employed to eliminate these artefacts on the 3D data: A spike removal filter is used to replace the outlier points; a hole-filling filter identifies and fills the holes on the face surface; a smoothing filter helps further remove the spikes and smooth the whole 3D model at last. However, the sequence and the choice of these basic filters vary for different implements, which might make significant impact on the final quality of 3D data.



(a) original (b) denoised

Figure 2.4: Different types of noise in a sample 3D face capture from FRGC database: (a) three different types of noise and (b) the corresponding denoised capture. The captures shown in (a) and (b) are plotted in the *xyz* space with different view angles with different ranges of *z* axis for demonstration.

The denoising method proposed in [39] is widely applied by 3D face recognition algorithms [40-42] and three operations that are mainly used to eliminate the artefacts are median filter, hole filling and smoothing. Specifically, the spikes on the face are removed by median filter and the holes caused by spikes removal are filled. Finally, the whole 3D face data is smoothed. Mian *et al.* proposed an effective denoising method: after face cropping, the outlier spikes are eliminated and holes are filled by the cubic interpolation. Finally, a median filter is employed to denoise the Z-component of the 3D face [5].

Compared to the methods used in [39-42], some other approaches [5, 43, 44] calculate and select the variance that is larger than a threshold value to remove the spikes instead of using the median filter. Mian *et al.* defined the outlier points according the distance to their 8 neighbouring points [5]. Chang *et al.* remove the spikes by calculating the variance in the Z coordinate within an 11×11 window around each points [43]. To remove the outlier vertices, Lei *et al.* smooth the values of all three coordinates (x, y and z) of the outlier vertices according to the statistical information of neighboring vertices [44].

Some other methods have been proposed to eliminate the spikes. In one of these, the observed points are defined as outliers if the angle between their optical axis and surface normal is larger than a threshold [45]. In [20], the outliers are removed during nose tip detection as the spikes are considered as the candidate nose tip points. Gaussian filtering is used in [46] to remove the spikes.

Generally, cropping the face region is the first step of denoising because it can remove the irrelevant details on the face such as the hair and neck [5, 20, 43, 45, 47, 48] while other methods crop the face after removing the noise [39-42, 44, 46, 49-53]. The order in which the filters are applied also varies among different methods. Spike removal, hole filling and smoothing filters are applied sequentially in [39-42, 51]. In [43, 44, 54], they remove and smooth the spikes before filling the holes and there is no further smoothing in the following steps.

Three major interpolation methods are used to fill the holes: cubic interpolation [5, 55], bicubic interpolation [44, 54] and linear interpolation [20, 43, 47, 49]. Some effective filling filters include the Gaussian filter [39-42, 44-46, 52, 54, 55], the median filter [5, 53, 56], the Wiener filter [20] and the mean filter [44, 47]. All of these are widely used for smoothing the range or point cloud data.

In addition, a number of other effective denoising methods have been proposed in the recent years. Three different Gaussian filters are employed to eliminate spikes, fill small holes and smooth the data [46]. K-nearest neighbour interpolation is applied to remove or smooth the noise [48]. Drira *et al*. sequentially utilise the VTK (www.vtk.org) library to fill holes, crop main the face region and smooth the spike [49].

2.2.4 Alignment

Alignment aims to remove the pose variances in the input 3D captures and normalize them to a predefined pose. It is an indispensable preprocessing step in particular if the features to be extracted in the following steps are sensitive to pose variations. Misaligned captures might lose the correspondence in feature space of the same subject and the recognition performance will be significantly reduced. Moreover, the alignment algorithms should be efficient to meet requirement of real-time automatic face recognition systems.

There are numerous algorithms that have been proposed for 3D face alignment, which can be categorized as template based, landmark based or shape based [57]. The Iterative Closest Point (ICP) algorithm [58] is the most well-known template based approach and iteratively changes the rotation and translation of the probe image to minimize the energy function. This energy function is defined by comparing the current state of the probe and gallery. Although ICP was originally proposed for 3D shape registration, it has been widely used for 3D face recognition. The main drawback of ICP based methods is that they might not produce good performance for coarse alignments as, being an iterative algorithm, ICP can be trapped in local minima. Annotated face model (AFM) [39] and adaptive active appearance models (AAM) [59] are also the examples of template based approaches. The main drawback of template based algorithms is high computation cost for alignment.

As the curvature calculation is rotational invariant [60], it is widely used in landmarking and then the detected pose invariant landmarks are applied to correct the rotations [37]. The major drawback of this approach is that the alignment performance is significantly affected by the accuracy of the landmark detection. In fact, landmarking is very challenging and the accuracy can be degraded by the noise and self-occlusion.

In comparison, shape based approaches do not rely on the reference captures used in the template based algorithms and are also less sensitive to specific landmarks. Shape based approaches utilise the regional or holistic facial geometry information to correct the pose

rotation. According to the symmetry characteristics of facial surface, both halves of the face are compared by ICP to correct the pose variation in [46]. Mian *et al*. proposed a 3D face alignment method based on PCA transform [5], which has been widely applied as a preprocessing procedure in many recognition algorithms [7, 34, 61]. It is a fully automatic algorithm for the pose correction of 3D captures. Compared to other existing techniques, this method only used the nose tip information and then iteratively corrects the pose using the Hotelling transform [62], which is also known as the PCA. The holes are filled by cubic interpolation and the face is resampled on a uniform square grid. The main problem of this PCA based alignment method is that the order of axes might suffer unwanted changes when the 3D points are mapped to the principal axes, which is difficult to detect.

2.2.5 Landmarking

In feature space manipulation, landmarking plays a significant role in region based face recognition algorithms. The accuracy of the landmarking will directly determine the effectiveness of feature extraction. A comprehensive summary of widely used landmarking algorithms in both academia and industry is provided by Creusot *et al*. [16]. Three categories of landmarking algorithms are used: curvature based descriptors, facial profile signatures and machine learning based methods. Using the curvature descriptor maps to find the salient points on the 3D captures is widely used by many researchers [4, 37, 38, 63-66]. For example, HK classification is the most popular method to detect the landmarks, which thresholds the mean (H) and Gaussian (K) curvatures to find the most salient candidates. In [4, 64], after the HK classification the landmark candidates are defined as the extrema points in given directions. However, the majority of algorithms utilising the HK classification are sensitive to yaw, pitch and roll rotations, even though H and K of a surface are rotation invariant [38]. In [37], a cropped face template containing three main landmarks (nose tip and eyes) is used for validation, which will also suffer variations under yaw and pitch rotations.

Instead of employing HK classification, the shape index (SI) and curvedness [35], computed from principle curvatures, can provide more effective maps for landmarking. In a typical and widely used method [63], the nose tip is detected by the largest SI and two inner eye corners are found by detecting the smallest SI. In most approaches, the nose tip is always located at the maximum of depth on the largest convex region.

Another popular approach is to find the facial profiles and transections of the 3D point cloud and use these to localize the nose tip, root and corners. For example, Faltemier *et al*. proposed the use of rotated profile signatures to extract the contours of the mesh at varying angles until it matches a previously learned nose profile signature [67]. Segundo *et al*. [66] and Mian *et al*. [68] have also used transverse slices to detect the nose tip and the nose corners. Instead of using a heuristic method for landmark detection, machine learning based approaches are applied in many methods [16, 69, 70]. Creusot *et al*. proposed a machine-learning based approach to build a strong feature vector from the 3D mesh by combining different local surface descriptors [16].

Emambakhsh *et al*. proposed an effective landmarking algorithm and found 16 nasal landmarks as shown in Figure 2.5(a), including the nose tip, root and two alar grooves [7], and successfully applied it to the Bosphorus [13] and FRGC [12] databases. Convex regions were found by thresholding the SI using $-1 < SI < -5/8$ [7, 27, 34] and the nose tip is located at the centroid on the largest convex region of the face. However, this method is not always robust for some scenarios: (1) the shape of the human nose is diverse and the nasal root is not necessary located at the minimum of each curve, which results in failed detection of some subjects; (2) denoising is still challenging in 3D face recognition and the algorithm is also sensitive to the residual noise; (3) as it is sensitive to the pose variations, accurate pose alignment is required and (4) occlusions by glasses.

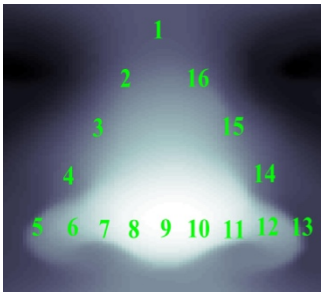


Figure 2.5: 16 nasal landmarks

2.2.6 Feature Extraction

Feature extraction is a crucial step in 3D face recognition and there are numerous approaches proposed for this step, which can be generally categorized into three groups, namely holistic, region based and hybrid-matching [71]. Holistic algorithms employ the whole face to recognise people, for example, Eigenfaces [72], Fisherfaces [73], Randomfaces [74], Gradientfaces [75], Laplacianfaces [76], Independent Component Analysis (ICA) [77], Support Vector Machines (SVM) [78] and Neural Networks [79]. Region based algorithms find some regions of interest,

including the nose, eyes and the mouth, and then attempt to extract features, which are relatively robust to variations on the face. Hybrid-matching algorithms combine holistic and region based methods to enhance the overall recognition performance.

Although many methods can be used for extracting features from the 3D face, depth information is commonly used because of its efficiency. One of the most popular approaches is to draw curves between extracted landmarks. For holistic algorithms, Drira *et al*. first manually located the nose tip and then used a set radial curves emanating from the nose tip to develop a Riemannian framework for analysing the shape of full facial surfaces [49]. Although this approach overcomes some of the significant problems in face recognition and produces a high recognition performance, utilizing information from the whole face region is challenging and also requires a non-trivial preprocessing step to normalize the whole face and correct for occlusions. Ballihi *et al*. extracted a set of geometric features including circular and radial curves [80]. However, the location of the curves covered some expression sensitive regions presented problems for expression robust recognition and the nose tip also had to be detected manually. Emambakhsh *et al*. automatically detected four significant landmarks, the nose tip, the saddle and the two alar, and used these to define a set of 28 nasal curves that were used for expression robust recognition [7].

Instead of using the depth information or point clouds, many other 3D representations can be used for feature extraction. These includes surface normals, that can be considered as a vector [17, 20] or three components [6, 61], the wavelet transform [39], multi-resolution wavelets [46], Local binary patterns (LBP) [81], the DCT [82], curvatures (principle, Gaussian, mean, shape index or curvedness) and Gabor filters [47, 83].

In [39], to improve the efficiency, a deformable model framework is used to calculate the differences between the facial scan and the model and then the resulting differences are converted to a 2D geometry image and transformed to the wavelet domain. This work proves that a small portion of the wavelet data is sufficient to accurately describe a 3D facial scan. Local features described by Gabor wavelets [47] are extracted from depth and intensity images to improve the whole recognition performance. Employing the Gabor wavelets, different local features on the facial surface can be found by varying scales and frequencies and those features from different components are fused to build a strong feature vector.

Spherical Face Representation (SFR) is utilized to quantize the 3D data by mapping the face into concentric spheres [5]. SFR and the 2D Scale Invariant Feature Transform (SIFT)

descriptor are combined to form the rejection classifier to narrow the search domain in the gallery. Combining with Haar, Gabor and LBP, Wang *et al*. introduced a method, termed the Collective Shape Difference Classifier (CSDC), to calculate the differences between the gallery and probe to make the feature space more various [46].

Instead of using the depth or points' coordinates for 3D registration, iterative closest normal points (normal vectors) are employed to find the corresponding points between a generic reference face and every input face [20]. The use of the point normals can maximize the within-class similarity and between-class distribution. In [17], the Eigenfaces, Nonnegative Matrix Factorization (NMF) [84] and ICA is also applied to the normal vector. All the captures are reconstructed from photometric stereo and this device is called Photoface. Li *et al*. utilize three different scales Local Normal Patterns that are in a similar manner of LBP to extract the facial features on three components of surface normal [6]. The resulting histograms from the nine facial normals are used to build the final feature space.

For region based algorithms, curvature calculation is always applied first to find the target region, such as convex and concave parts, and then other methods used to extract the features. In [85], curvature information is applied in face segmentation and landmarking and 7 overlapping regions are used for recognition. In addition to the point cloud, curvature based 3D shape descriptors are used to build the feature space.

An alternative approach is to use relatively expression robust regions or structures for feature extraction in recognition systems [86]. Lemaire *et al*. divided the face into several regions on the basis of Facial Action Coding System (FACS) and anatomic considerations [87]. Patches around the landmarks on the face region are found by fitting of the Statistical Facial Feature Model. Then the patches were compared with six prototypical facial expressions using ICP and the distance between the patches was used as features for classification. Patches were found around landmarks in the 3D mesh, which is defined by circling the points [88]. The length of the geodesic path between corresponding patches, computed using a Riemannian Framework, provides quantitative information about their similarities. These measures are then used as inputs to several classification methods

To address the pose variations, some rotation invariant feature extraction algorithms are widely employed, for example, 3D Fourier descriptors [89], Extended Gaussian Images (EGI) [90], spin images [91].

2.2.7 Postprocessing

The feature sets extracted from 3D captures always possess high dimensionality, which will significantly increase the storage costs and computational complexity. To reduce the dimensionality and find the most effective features, feature selection algorithms can be applied. Those are generally categorized into two groups: deterministic sequential and stochastic search based methods. The sequential methods search for the feature combinations in a predefined order. Forward sequential feature selection (FSFS) is a famous deterministic algorithm proposed by Pudil *et al.* [92], which is a low cost and effective feature selection method [93]. It is a top-down search strategy and iteratively concatenates the feature sets to obtain the highest recognition performance. In contrast, backward sequential feature selection (BSFS) initially utilizes the whole feature space and iteratively removes the feature sets [93].

Stochastic search based methods aim to maximize the objective function by evaluating various combinations of feature sets and selecting those which possess high probability to achieve the global optimum. To overcome the drawbacks of the traditional gradient descent based method, genetic algorithms (GA) and simulated annealing are used as non-deterministic search based feature selection approaches [94]. GA, as evolutionary algorithms, are one of the widely used heuristic algorithms, whose evolution usually starts from a population of randomly generated individuals with an iterative process. GA have the capability to examine various combinations of the features. However, convergence is not guaranteed to occur in a limited number of iterations. Compared to sequential algorithms, search based approaches possess higher computational complexity.

Instead of using feature selection strategies, the reduction of feature space's dimensionality and feature vectors' correlation is another effective postprocessing method to achieve a good classification performance as it is not necessary to use high dimensional feature sets [95, 96]. PCA is the most widely used unsupervised method for dimensionality reduction and Linear Discriminant Analysis (LDA) projects the data to a lower dimensional space to maximize the within-class similarity and between-class dissimilarity. For linear Fisher's analysis, PCA is applied before LDA implementation. However, the main problem of PCA based methods is that using linear transformation will not necessarily find the maximal variance direction of features' distribution, as some classes might not be linearly projected. To address this problem, non-linear projections are proposed to effectively distribute the features. Kernel PCA (KPCA), Kernel LDA (KLDA) and Kernel Fisher's Analysis (KFA) are widely used kernels based

approaches while linear, polynomial, quadratic, and radial basis functions are the most typical kernels.

2.2.8 Matching and Classification

For each probe (p), calculating the distance from the gallery (g) is the simplest way to produce the matching scores. In Table 2.1, four widely used distance calculation methods are listed, where ◦ is the componentwise Hadamard product operator, Σ(.) computes the sum of the elements of a given vector and σ_m is the covariance matrix computed using the gallery samples.

There are numerous matching strategies explained in the literature. Boosting is an effective algorithm for classification. Instead of building a strong classifier, boosting algorithms aim to iteratively merging weaker classifier to produce higher recognition performance. It can be used for face recognition [8, 80, 97, 98]. Hidden Markov models (HMM) are proposed for recognizing range images of faces [99]. In [100], embedded HMM are also used for both 2D and 3D face recognition algorithm.

In addition to the facial alignment, the ICP algorithm can also be applied in computing the matching score [101]. Also, the distance between the detected landmarks are obtained in [102], and dynamic programming and Support Vector Machine are applied for classification. The Hausdorff distance [103] is used as the matching criterion for both 3D and 2D face recognition problems [104].

Sparse representations are proposed for classification [74]. The main drawback is that the classification performance might degrade when there are insufficient training samples for each class and is also sensitive to facial variations. Deng *et al.* [105] proposed to learn occlusions and artificially create more samples to each subject. Compared to the traditional applications, a weighted sparse representation is proposed for face recognition [6].

Table 2.1: Four widely used methods to calculate the matching score between the probe and gallery.

Euclidean distance:	$D = \sqrt{\sum (p-g) \circ (p-g)}$
City-block distance:	$D = \sum \lvert p-g \rvert$
Cosine:	$D(\text{in radians}) = \cos^{-1} \frac{\lvert p \cdot g \rvert}{\lvert p \rvert \lvert g \rvert}$
Mahalanobis distance:	$D = p\sigma_m g$

2.2.9 Neural Networks and Deep Learning

The traditional hierarchy approaches that are discussed above are widely used to solve the pattern region problems and are known as 'shallow' methods [106]. Many classic 3D face recognition strategies suffer from the generalisation problem. 3D captures from different data capture devices may possess various characteristics. For example, the data captured from laser scanners has high quality and preserves better facial details. In comparison, facial information obtained from photometric stereo is likely to suffer problems caused by integration during the reconstruction and the captures are also quite smoothed which might lose some fine details. Therefore, feature extraction algorithms might have various implementations or development for different kinds of data, which is hard to design a generalised algorithm to meet all the types.

Neural networks are proposed to address this problem and in [107] a literature review on its applications is provided in terms of Probabilistic Neural Network [108], Radial Basis Functions neural networks [109], and convolutional neural networks [110]. However, the traditional algorithms suffer high computational cost for training. In recent years, with the development of GPU's parallel and cloud computing, deep learning algorithms are being applied to face recognition, aiming to address the shortcomings of the shallow methods. Compared to the shallow algorithms, the deep learning architecture performs feature extraction, selection and classification in each trainable unit. The deep learning strategy is similar to the way of automatic feature extraction and selection in the human brain which uses different neuron layers to train the features.

However, the drawbacks of such kind of approaches are also apparent and there are still some ongoing research or application problems. In general, the computation speed of deep learning algorithms is much slower than that of shallow algorithms. Using an inappropriate number of neurons may result in the over or under fitting problems and it is often hard to find specific parameters of layers to produce high training results. Finally, some deep learning architectures can only achieve good recognition performance on some specific datasets.

2.3 3D Face Recognition under Expressions

The human face, regardless of whether is a 2D or 3D representation, has high probability of exhibiting different kinds of expressions. It is hard to constrain all the facial captures to the neutral or the same expression in realistic biometric application scenarios. In fact, compared to other variations caused by pose or occlusions, facial expressions result in facial muscle

movements that may significantly deform the face surface. This causes problems for feature space manipulation and makes the subject lose the within-class similarity and between-class scatter. Therefore, matching images with different expressions in the probe and gallery is one of the most challenging problems in 3D face recognition and the development of recognition techniques that are relatively robust to expression variations has been the subject of much research over the past decades.

Many 3D face recognition algorithms are proposed to address the expression variations and Smeets *et al*. provide a comparative study of 3D face recognition under expression variations [111]. In general, expression robust 3D face recognition algorithms can be categorized into two groups in terms of rigid and non-rigid strategies. Rigid approaches aim to investigate the constant facial patches under expressions, for example of the 3D nasal region. The forehead is also considered as the relatively expression robust region [5], even though it suffers more hair occlusions. In contrast, the non-rigid strategy, attempts to transform the facial captures with expressions to the neutral face by a deformable model.

For the rigid strategy, one effective method is selecting relatively stable structures on the face for expression robust discriminative features extraction. Wang *et al.* proposed to use shape difference boosting for evaluating the deformations caused by expressions in facial patches and select some stable facial region under expressions [8]. The Bosphorus database is used for facial patches selection and the FRGC v2.0 database is used for recognition performance evaluation. In a similar manner, Li *et al.* divide the whole face into different scales of rectangular patches and select the most expression robust patches using a weighted sparse representation [6]. The features extracted from each patch are obtained from the depth and three components of the surface normals. Mian *et al*. exploit the recognition performance of the whole face and show that the nasal region and forehead are the most robust structures under expressions [5].

Instead of selecting facial patches, the nasal region, which is considered to be a relatively rigid region on the human face, has been widely utilized in 3D face recognition in the recent years. The nasal region has been shown to be more consistent over natural expressions and occlusions. Chang *et al*. matched multiple overlapping 3D nose and its surroundings and obtained a good recognition performance [4]. Wang *et al*. also explored different size nasal and its surrounding region by changing the radius of a sphere centred on the nose tip, which indicated that the performance of nasal region is equal to that of the whole face [8]. Ballihi *et al*. found that

circular curves around the nasal region produce a better recognition performance than other curves [80]. All the results show the potential for employing nasal region to find discriminative features and its significant contribution to face recognition. Some algorithms that extract features on the nasal region for expression invariant face recognition are discussed in Section 2.4.1.

Using a non-rigid strategy, Kakadiaris *et al*. employed the deformable model framework and pyramid transformation to address expression variations [39]. Using the deformable model framework, all the 3D facial captures are accurately aligned to fit the model, even in the presence of facial expressions. The 3D difference between the input capture and facial model is saved as a 2D geometry image, which is further transformed to the wavelet domain. The pyramid transformation is applied to solve the position and rotation changes caused by expressions, as the pyramid wavelet is translation and rotation invariant. To make the features less sensitive to expressions, it was proposed to apply the ICP algorithm to the surface normal vectors instead of depth representation, for face registration and find the corresponding points between a generic reference face and each input face [20].

Inspired by the introduction of sparse representation [74] for face recognition, Deng *et al*. proposed to address the problem of expression variations by artificially creating expressions for each capture [105]. The additional expression variations, as well as other variations in terms of illuminations and occlusions, are learnt from the given database. The newly added captures increase the number of training samples of each subject, which is beneficial to the classification performance of sparse representation as the original proposed method fails if training samples are insufficient.

2.4 Using the Nasal and Adjoining Regions for Recognition

2.4.1 The Motivation of Using 3D Nose

Some relatively stable structures and patches on the face can be used to help design expression invariant face recognition algorithms. Mian *et al*. evaluated some regions on the face and found that the forehead and nose are more consistent over different expressions [5]. This study proved that the facial movement caused by natural expressions is not salient on those regions. However, the forehead is sensitive to the occlusion by human hair which might present problems for feature extraction. In comparison, the nose is very difficult to deliberately occlude by the hair, hands and scarves [112].

The nasal region has not been widely used for region based 2D face recognition algorithms, as it was believed that it would not possess enough discriminative features when compared to other parts in 2D texture images [18]. The introduction of depth information has enabled more nasal shape properties to be found and explored, and this has increased the potential to develop nose recognition algorithms over the past decade. Using the 3D nasal region as an additional biometrics demonstrates high accuracy and robustness in comparison with using the 2D nose. In addition, using the 3D nose region can provide a better solution than exploiting the whole face, in terms of expression, pose and occlusion variations.

Moreover, using the curvature information and facial profile signature, it is relatively straightforward to detect and segment the central region of the 3D face that contains the nose. There are some salient landmarks on the nasal region such as the nasal tip, root and alar grooves, which can be robustly and accurately detected by their shape properties. For example, the nose tip is usually the closest point to the camera and its convexity is more salient than other parts of the face. Using these landmarks, some discriminative features, including curvature, depth/surface normal differences and curves, can be extracted from the nasal region, which offer many advantages for expression robust 3D face recognition.

2.4.2 Overview of 3D Nose Recognition Algorithm

There are many advantages mentioned above of using the 3D human nose as a biometrics and some researchers have focused specifically on 3D nose recognition or have evaluated the recognition performance of the nasal region. Chang *et al*. matched multiple overlapping 3D regions containing the nose and its surroundings and obtained good recognition performances, thus demonstrating that this region has much potential for providing expression robust biometric features [4]. In the preprocessing phase, the face is detected by 2D colour information and aligned by ICP. Some significant landmarks including the nose tip, eye corners and the nasal bridge are localized by HK classification. The recognition is performed by matching the nasal region of both the probes and gallery and the performance is evaluated using the FRGC database, Spring2003 folder as the gallery and the other two folders as the probes. The probes are divided into neutral and non-neutral groups for demonstrating the influence of human expressions. Generally, the expression variations might cause about 10% recognition performance reduction in most applications.

Dibeklioğlu *et al*. propose a fully automatic pose and expression invariant 3D nose recognition system [113]. The main contribution of this work is the segmentation and pose calibration of the 3D nose. The nose tip is first detected by HK classification and shape index thresholding and then a coarse to fine nose segmentation and pose estimation and correction algorithm is applied to extract the nasal region. The ICP algorithm is used for recognition and the performance is evaluated on the Bosphorus database, which contains a lot of captures with various expressions or ground-truth pose variations. The algorithm achieves a R_1RR of 94.10% and 79.41% for frontal and rotated face captures, respectively.

Wang *et al*. also explored the use of nose and its surrounding regions to build a more efficient classifier by shape difference boosting on different cropped region of the face [8]. To crop the nasal region, different radius spheres centered on the nose tip are used and it was found that the recognition performance improves when more regions are included in the matching. For a 24 *mm* radius sphere, that mainly includes the nasal region and a 44 *mm* radius that contains the nose and some part of cheeks and mouth, a R_1RR of 77.5% and 95%, respectively, is achieved.

Drira *et al.* introduced a geometric analysis of 3D nose shape under different facial expressions and demonstrated the nasal region contribution in 3D face biometrics [114]. To quantify the shape differences, the length of a geodesic between any two arbitrary nasal surfaces is estimated as the geodesic length between a set of their nasal curves. This method was tested on a small subset of the FRGC database, including 125 neutral gallery images and 125 non-neutral probes. The recognition performance outperforms the baseline ICP algorithm, which demonstrates the potential of using the nasal region for expression robust 3D face recognition.

Moorhouse *et al.* proposed a 3D nose recognition algorithm, which is mainly focused on extracting discriminative features on the reconstructed 3D photometric stereo captures [3]. HK classification is used for landmarking and nasal region segmentation. Geometric ratios, Fourier descriptors of the nasal ridge, a combination of the features and Eigennose are employed on the nasal surface to build the feature space. Compared to the previous public work, results are presented for a small subset of the Photoface database. Although this is one of the first 3D nose recognition work tested on this kind of reconstructed captures, the recognition performance is relatively low and the highest R_1RR obtained is around 47% by the Eigennose algorithm.

Emambakhsh *et al*. proposed an expression robust 3D nose recognition algorithm by matching nasal curves [7]. After the denoising, face region cropping and pose alignment, 16 landmarks

are robustly localized on the nasal region. A set of 75 curves are then defined by connecting selected pairs of landmarks on the cropped nasal surface. FSFS and GA are employed to select the most discriminative curves and a total of 28 curves are used to build the feature space. A novel landmarking strategy is proposed to find four main landmarks: the nose tip, root and two alar grooves. The Bosphorus database is used for feature selection and the extracted features are further tested on the FRGC v2.0 database.

2.4.3 Using the Adjoining Cheek Regions for Recognition

Although the cheek region has been shown to be more susceptible to expression variations than the eyes, forehead and the nose [5], the adjoining regions between the cheek bone and nasal bridge have the potential to producing rich discriminative features, which can be used as part of an expression robust recognition scheme. Chang *et al*. produced a good recognition performance by matching multiple overlapping 3D region of the nose and its surroundings [4], demonstrating that such regions have much potential for improving expression robustness. Wang *et al*. also explored the nose and its surrounding region to build a more efficient classifier [8]. Ballihi *et al*. found that the most relevant circular curves are located around the nasal region and its surrounding regions [80]. Therefore, extending the nasal region to include adjacent areas is capable of achieving a high classification performance for expression robust 3D face recognition.

Recent 3D face recognition work using the nose region has drawn 28 curves joining 16 robustly detected landmarks on the nose [7]. However, this work only considered landmarks and curves directly on the nasal surface and the regions adjoining the nose, between the cheek bones and nasal bridge, were not included. In fact, these regions are also relatively stable and less affected by occlusions when compared to other patches on the face surface. These regions might contain additional discriminative features to those on the nose, which shows the potential of extending the nasal region to include adjoining parts of the cheeks for expression robust 3D face recognition.

2.4.4 Expression Robustness Evaluations on Various Facial Patches

In this section, three different evaluations are proposed to investigate the recognition performance of different facial patches, providing the motivation to extract the discriminative features for 3D nose identification. The first of these is to calculate the within-class discrepancy

over many kinds of expression to demonstrate that the nose region is relatively rigid and its structure is more stable when expression variations occur on the facial surface.

The other two parts concentrate on evaluating the recognition performance using both large and small scale patches on the human face. This also proves that the nose region outperforms other facial parts and suffers from very few natural occlusions. Both the depth and surface normal maps are considered in the following three evaluations and results presented for all the captures in the Bosphorus database excluding those with occlusions and large pose variations.

2.4.4.1 Within-class Dissimilarity under Different Expressions

Dissimilarity maps were computed for globally registered faces using the point set features [33]. The maps were built by subtracting the captures with expressions from neutral captures of the same subject. Although these dissimilarity maps provide a good representation of the whole face and prove that the human nose is relatively stable under different expressions, only depth information was considered. Li *et al*. [6] used three components of surface normal (SN_x, SN_y and SN_z) calculated from 3D point clouds and explored the expression invariant discriminative features for recognition, demonstrating the potential of using surface normals. Therefore, three components of surface normals are also considered in this experiment. As a preprocessing step, the pose variations are first corrected and all the captures are translated so that they are centred on the nose tip.

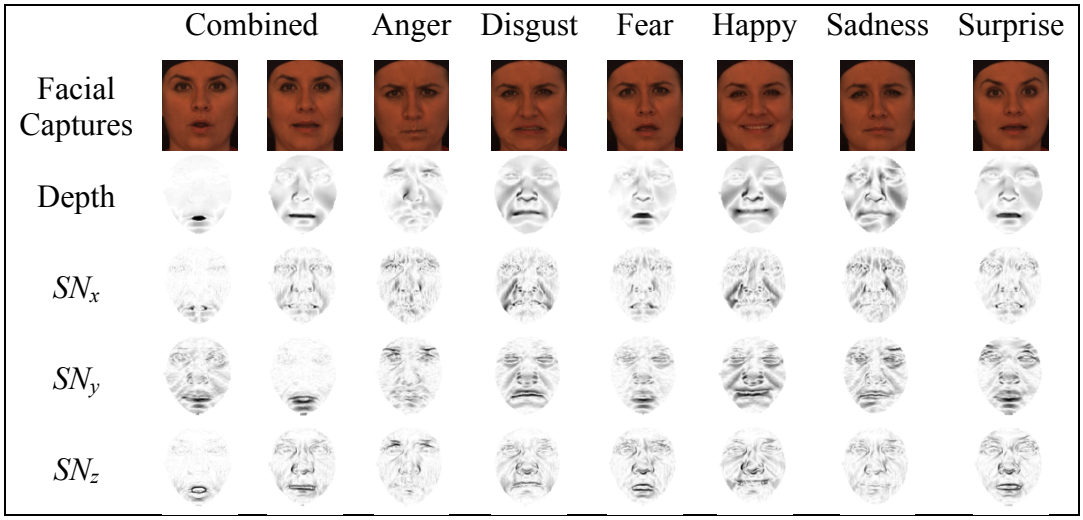
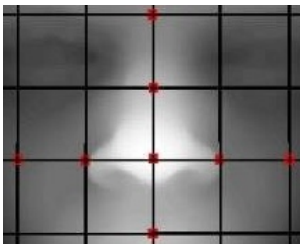


Figure 2.6: Dissimilarity maps calculated from the captures with different expressions and the neutral one using four components of 3D data on the Bosphorus database. Darker regions show greater dissimilarity on the face. Combined expression contains facial surface deformations on the upper and lower units.

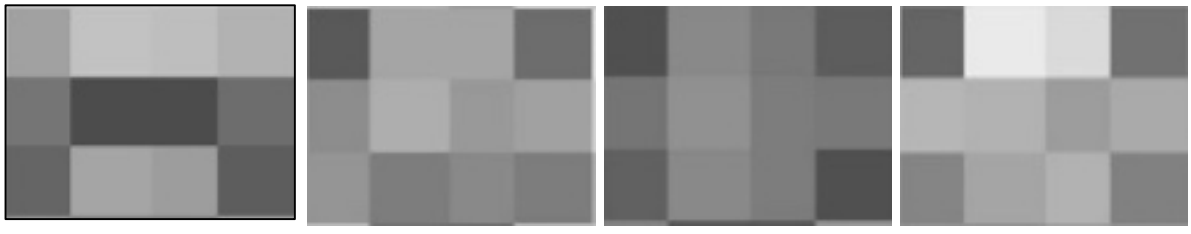
The Bosphorus database [13] provides a good choice for the expression invariant feature extraction, which contains Face Action Units that describe the facial surface changes when expressions occur on different parts of face (lower, upper and combined). It also contains some basic human expressions, including anger, disgust, fear, happiness, sadness and surprise. All captures with expression variations are considered in this evaluation.

To be more specific, all the Face Action Units are used for calculation which results in 32 dissimilarity maps of each component (depth, SN_x, SN_y and SN_z), including 2 combined units, 20 lower units, 4 upper units and 6 basic expressions. To demonstrate the changes on the face, 2 combined units and 6 basic expressions are illustrated in Figure 2.6. As can be seen from the maps, some small patches on the face show different variance in each component under specific expressions. For example, the cheek bone part is widely regarded as a non-rigid region which suffers more changes under expressions [5]. The depth maps shown in Figure 2.6 indicate that the cheek bone region has limited stability. However, the surface normals calculated on the cheek are more consistent, which motivates the investigation of different types of discriminative features extracted from the non-rigid regions. In general, as can be seen from Figure 2.6, the nasal region demonstrates higher stability under various expressions.

2.4.4.2 Large Scale Patches Evaluation Using Selected Landmarks



(a) 12 patches



(b) depth (c) SN_x (d) SN_y (e) SN_z

Figure 2.7: Landmarks based large scale patches evaluation on the depth and three components of surface normals. The brighter patches denote higher recognition performance. In general, features extracted from nasal region outperform other patches, which is more salient in the surface normals maps.

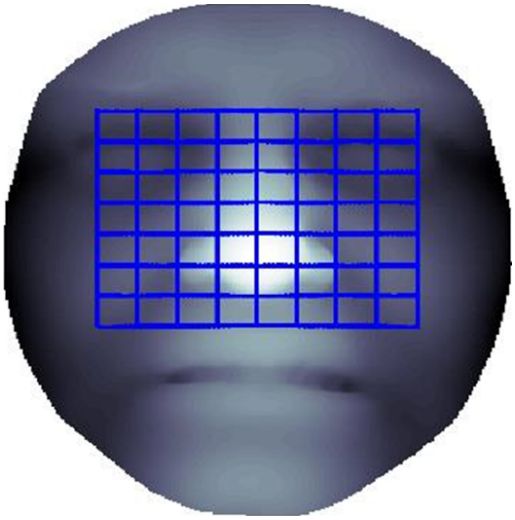
In previous studies, both 2D and 3D facial data are usually divided into fixed sized patches and for their recognition performance evaluation, LBP [115] is the most popular descriptor to extract features on each patch [6, 51, 116, 117]. For example, in [6] all the captures are first resampled to a fixed size and different scales of patches are used for recognition performance evaluation. In addition, the three components of the surface normals are also used to calculate the dissimilarity maps.

However, the main problem of these methods is that different subjects or their captures may contain varying content in the target facial patch. The underlying reason for this is that human faces possess their own characteristics (e.g. the size and curvatures) and their structure and distribution might be varying. Although such discrepancies can preserve the within-class similarity, they can have a great influence on the between-class dissimilarity. Therefore, in this section, an improved evaluation strategy to correct the content discrepancy is proposed. In Figure 2.7(a), using the nose tip as a reference seven further landmarks are automatically detected: (1) the nose tip, root and two alar grooves [7]; (2) two cheek landmarks; (3) middle nose bridge (middle point between nose root and tip) and middle subnasal (symmetrical to middle nose bridge). On the basis of these landmarks, 12 patches are found on the central facial region and each patch is resampled to a fixed size.

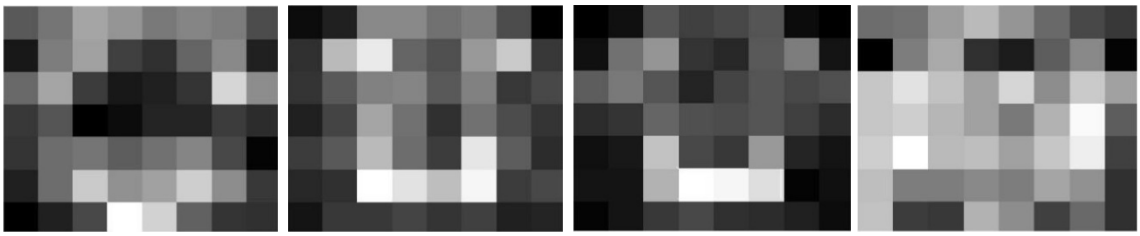
A set of LBP values is calculated for each patch and the resulting LBP histogram is used to build the final feature set. The recognition performance of each patch tested under identification scenarios is shown in Figure 2.7(b)-(e), where the brighter regions indicate a higher recognition performance. Compared to the other facial parts, the nasal and adjoining regions are more discriminative and has more potential to produce a good recognition performance.

2.4.4.3 Local Patches Evaluation on the Main Part of the Human Face

In addition to the large scale patches for the central face evaluation, 56 local patches, as shown in Figure 2.8(a), are used to further evaluate the discriminatory power on the nasal and surrounding regions. The local shape difference descriptor is used to extract features in each patch, which will be further explained in Section 5.4.1. The extracted features of each patch are evaluated in identification scenarios and the resulting discriminatory maps of four components are shown in Figure 2.8(b)-(e).



(a) 56 local patches



(b) depth (c) SN_x (d) SN_y (e) SN_z

Figure 2.8: The discriminatory power maps of local patches on the main part of the human face. The brighter patches denote higher recognition performance. Similar to the larger patches evaluation, patches on the nasal and adjoining regions produce better recognition performance.

As before, the brighter regions indicate a good recognition performance. Compared to other patches, patches from the nasal region generally perform better than those on the depth, SN_x and SN_y maps, especially for the lower nasal part. For the SN_z map, the nasal region produces a better recognition performance than the eye and upper mouth regions but worse than the adjoining cheek region.

2.5 Conclusions and Ongoing Challenges

The importance of using various biometric modalities is first introduced and the motivation of using the 3D face as a biometric is further explained in Section 2.1. For providing an overview of face recognition algorithms, some key techniques proposed and widely used in literatures are also summarised in this chapter. To address the expression variations on human face, an introduction and the motivation of using nasal and adjoining regions for 3D face recognition is provided in Section 2.3 and 2.4. There are some ongoing challenges in 3D face recognition literatures:

The nasal and adjoining cheek regions demonstrate great potentials to produce good recognition performance when face captures contain expressions. However, how to select an appropriate descriptor to extract discriminative and robust features is challenging in feature space manipulation. Novel descriptor which can better represent the local nasal structure is encouraged in the future work.

Also, many expression robust 3D face recognition algorithms only utilize some 3D face databases that contain the artificial expressions. Some of them are even exaggerated or unreal expressions. Although those types of expressions cause more facial surface movements and are more challenging in feature extraction, the naturalistic variations are still the most common expressions in the real world applications, which should be taken into consideration and evaluated in the future work. Photoface database offers access to naturalistic expression variations captured from a working biometric system in the field.

The Photoface device facilitates the 3D data acquisition procedure and provide a good overall reconstruction in comparison with the "ground-truthed" captures. The resulting surface normals possess higher resolution than other commercial acquisition system. Actually, there have been limited 3D face recognition algorithms tested or developed for Photoface database, even though the usage of Photoface captures can significant improve the real world application possibility. The region based feature extraction methods are hard to apply to the Photoface capture without effective and accurate landmarking strategies. It is interesting to apply the well-designed face recognition algorithms to the Photoface captures or investigate new techniques on the basis of their characteristics.

Curvature based landmarking is sensitive to the facial rotations and the nose tip is not always robustly detected for some captures. One simple and effective way to address those problems is to add differential geometry constraints to optimize the location of candidate landmarks. As a first order surface differential quantity, surface normals (SN_x, SN_y and SN_z) demonstrate the orientation of a surface and provide an effective way to localise landmarks.

Chapter 3 Databases and Preprocessing

3.1 Introduction

The selection of a suitable database is a vital component of 3D face recognition performance evaluation. The increasing number of databases that have been developed over the past decades has considerably helped researchers in this aspect. There is a growing need to consider databases' different characteristics to select the most appropriate one to use. Therefore, an overview of 3D face database is first introduced in this chapter.

The three databases, FRGC, Bosphorus and Photoface, that are used in the following chapters are further explained and their experimental settings are provided in Section 3.3. On the basis of varying characteristics in 3D face databases, the basic techniques discussed in Section 2.2 might have different application strategies. For example, 3D captures obtained from certain acquisition systems might suffer from the noise or the lack of face detection, cropping and alignment, hence requiring more preprocessing work to address them. Meanwhile, using different kinds of data as shown in Figure 1.1, the facial shape might demonstrate different properties. As a consequence, for real world applications, feature extraction methods investigated on different facial data are beneficial, as existing well-designed recognition algorithms may not necessarily achieve good performance on captures from different 3D face acquisition systems.

3.2 An Overview of 3D Face Databases

Although 3D face recognition algorithms have the potential to address problems in 2D approaches and achieve high recognition performance, there are still many other important aspects of 3D approaches that should be considered. One of these is the consideration of databases' characteristics used for recognition performance evaluations. As shown in Figure 3.1, the databases' characteristics considered in this chapter can be grouped into four main categories: Database Population, 3D Data Acquisition, Pre-processing and Landmarks, and Modes of Variation.

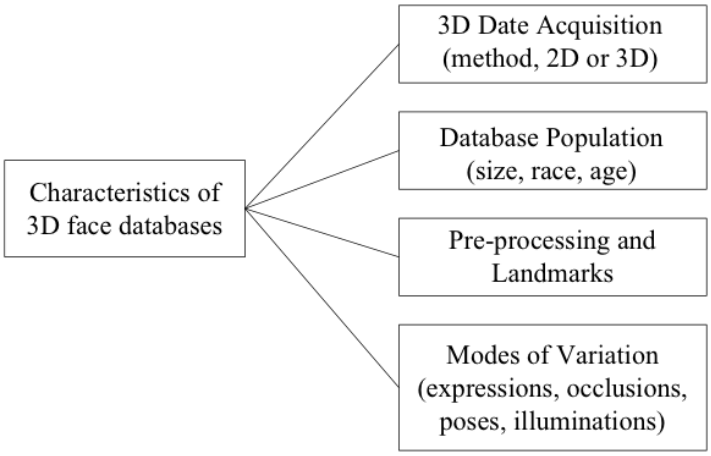


Figure 3.1: Classification criteria for 3D face databases.

3D Data Acquisition – The 3D capture technology used has a great influence on the data properties such as the resolution and noise characteristics, which can affect the preprocessing procedure and the accuracy of discriminative features extracted for recognition. For example, the 3D data obtained from some devices may contain more accurate representation of the human face but with more noise. Reconstructed captures may possess less noise but may suffer from the distortion of facial data during the reconstruction phrase. Therefore, on the basis of different application requirement, researchers need to consider the properties of different data acquisition as it is an important feature of 3D face database evaluation and recognition algorithm development.

Database Population Scope– A database's size is a key factor in determining its usefulness. Both the number of subjects and the total number of captures are important, as well as their gender and ethnicity. As the captures obtained in different periods of the same identity are advantageous for research, age is also an important parameter for consideration. A recent development has considered the genetic relationship between subjects, thus siblings can be found in several databases.

Pre-processing and Landmarks – Some databases include data/algorithms that help researchers in the production of the feature vectors that are used in the matching stage. Typically, face detection and segmentation may be included while landmarks often form the initial step of a face recognition algorithm. For example, the whole face region or nose region can be cropped using the nose tip and other landmarks can be used to find specific patches or curves on the 3D face surface. As a consequence, many recognition techniques require landmarks and some databases provide these as metadata.

Modes of Variation – The performance of many face recognition algorithms can be degraded by large facial variations including the pose, subtle skin movements by expressions, occlusions and lighting changes with varying posture [118]. As the development of recognition algorithms that are robust to variations is very desirable, databases that contain various facial variations can enable the robustness evaluations and at the same time more challenging captures provided in those databases can help contribute the development of face recognition algorithms.

3.2.1 Basic Characteristics of 13 Widely Used 3D Face Databases

In the last decade many 3D face databases have been collected by the research community. Table 3.1 presents, in chronological order, 13 publicly available databases including both true and reconstructed 3D data and gives details on their populations and the number of citations on Google Scholar on 13th February 2016. The first ten databases introduced represent the true 3D captures group and the last three are the reconstructed group. The remainder of this section briefly discusses the background and development of these databases.

Charles and Marc proposed the 3D_RMA database [119]. The 3D Data is captured by a prototype of a proprietary system based on the structured light. Although the quality of the 3D data is relatively low, it is sufficient to show the potential of 3D face recognition.

The FRGC database has had a large impact on the development and evaluation of face recognition algorithms and consists of 2D and 3D captures partitioned into training (v.1) and validation (v.2) sets, captured under both controlled and uncontrolled lighting conditions and with varying expressions [12]. Since its publication, it has been become a standard 3D face database to evaluate recognition algorithms. ND-2006, as a superset of the FRGC v.2, is the largest database available to the research community [120].

The GavabDB and FRAV3D databases were both developed by Face Recognition and Artificial Vision Group of Universidad Rey Juan Carlos [121, 122]. In particular, FRAV3D is a multimodal (2D, 2.5D and 3D) face database with 16 captures per subject [122].

The CASIA database was developed by the Chinese Academy of Sciences and considers combinations of different expressions, poses and illumination conditions, in addition to individual variations of each of these [97]. Another famous large-scale Chinese 3D face database, BJUT-3D, was developed by Beijing University of Technology and contains 500 Chinese people without expressions and occlusions [123].

The Bosphorus database can be employed in different aspects of face recognition research as it contains a rich set of expressions (35 Action Units as well as the six basic emotions), systematic variation of ground-truthed poses and different types of occlusions [13]. The second stage of the University of York 3D face database (York2) uses uncontrolled lighting conditions during the data acquisition and face captures have a variety of expressions and poses [124].

BU-3DFE is a rich 3D facial expression database which includes both 3D facial expression shapes and 2D facial textures. It can therefore be used for both expression robust 3D face recognition and 3D face expression recognition. Four intensity levels of six different emotions are captured for each subject [125].

The Texas 3DFRD [126] is complementary to other existing 3D facial databases (FRGC, Bosphorus and BU-3DFE) and uses a high resolution stereo imaging system. Furthermore, its colour and range images are perfectly registered to each other as they are captured simultaneously. UMB-DB contains multimodal (3D+2D colour images) facial acquisitions with a particular focus on facial occlusions [127]. In total, there are 590 captures with occlusions.

Table 3.1: Some characteristics of 13 widely used and publically available 3D face databases.

Databases	No. Subjects	Total No. Captures/ Captures per subject	Most Age Range	Texture	Race	No. Cites	Institute
3DRMA [119]	120	720/6	20-60	No	Caucasian	153	Royal Military Academy
FRGC v.2 [12]	466	4007/1-22	18+	Yes	White, Asian and others.	1846	University of Notre Dame
ND-2006 [120]	888	13450/1-63	?	Yes	Caucasian	42	University of Notre Dame
GavabDB [121]	61	549/9	18-40	No	Caucasian	171	Universidad Rey Juan Carlos
FRAV3D [122]	106	1696/16	18-35	Yes	Caucasian	22	Universidad Rey Juan Carlos
BU-3DFE [125]	100	2500/25	18-70	Yes	White, Black, East-Asian, Middle-east Asian, Hispanic Latino and others.	527	Binghamton University
CASIA [97]	123	4624/37-38	?	Yes	Chinese	14	Chinese Academy of Sciences
BJUT-3D [123]	500	500/1	16-49	Yes	Chinese	9	Beijing University of Technology
York2 [124]	350	5250/15	?	Yes	Various races	46	University of York
Bosphorus [13]	105	4666/31-53	25-35	Yes	Caucasian	323	Boğaziçi University
Texas 3DFRD [126]	118	1149/1-89	22-75	Yes	Caucasian, African, Asian, East, Indian and Hispanic.	85	University of Texas
UMB-DB [127]	143	1473/9-12	19-50	Yes	Caucasian	32	University of Milano Bicocca
Photoface [9, 17]	453	3187/?	?	Yes	Caucasian	17	University of the West of England

The Photoface database was captured in more realistic environment and also provides the detailed metadata for its capture [9, 17]. Each capture comprises of four differently illuminated images of the subject, from which surface normals, albedo and depth images can be calculated. It is a new reconstructed 3D face database but has more potential to achieve good recognition performance.

3.2.2 3D Data Acquisition

Capturing the 3D data in a biometric session is generally the first main step of a 3D face recognition algorithm and much original and significant information can be obtained during the 3D data capture process. One of main advantages of 3D face recognition algorithms is that, while 2D images are likely to be affected by the illumination in various ways, 3D representations can provide the illumination independent data [28-31]. Therefore, invariance to illumination is a significant contributory factor behind the enhanced performance of 3D face recognition techniques over their 2D counterparts. However, for many scenarios the process of capturing the 3D model is not illumination independent, as the acquisition of 3D shape by either the stereo or structured light involves one or more standard 2D intensity images. Similar 3D data acquisition problems may occur for any of 3D sensors currently used by 3D databases.

Some 3D sensors take various approaches to cope with these problems. Different 3D sensors and types of 3D data that are widely used by 3D databases are shown in Table 3.2 and Table 3.3, respectively. 3D face databases can be generally grouped into two distinct families: “True” 3D scans use 3D sensor systems such as the Cyberware or Minolta that can generate high resolution 3D captures. Alternatively, 2D image databases with multiple 2D cameras or illuminations allow for 3D model generation using techniques such as the photometric stereo or multi-view stereo reconstruction to recover an approximation of the 3D structure from 2D images.

Structured light is one of the most widely used methods for 3D facial data acquisition. It projects one or more encoded light patterns onto the scene and then measures the deformation on the objects’ surfaces to extract shape information. Its advantages include relatively low cost and the fact that it can be widely employed for real-time simultaneous acquisition of sequences of 3D facial surface and 2D intensity. However, it only provides lower spatial resolution images, resulting in samples that appear to have missing parts and holes [9]. Many acquisition systems are based on structured light technologies such as the Minolta Vivid 900/910 series (Konika Minolta Holdings, Inc., Tokyo, Japan) and the Inspeck Mega Capturor II 3D.

3DRMA [119] is one of the first publicly available datasets containing 3D face samples collected using structured light. The Bosphorus [13] and York2 [124] databases also employ structured light during their data acquisition but use more sophisticated commercial technologies [119].

The laser scanner uses a laser stripe projector and triangulation to build a 3D model consisting of a range image and a very accurately registered colour image for each session. The benefit of laser scanners is that they offer extremely high accuracy. However, during the reconstruction phase, some factors (e.g. specularities) may cause considerable artefacts which are manifested as spikes and holes on the surface. Therefore, as the sensor responds to a very narrow range of the illumination, the quality of the captured 3D shape can be degraded by relatively small variations in lighting [27]. In addition, full cooperation of the clients is required as there is a 2.5 seconds delay between the captures of the range and texture [128]. FRGC v.2 [12], ND-2006 [120], CASIA [97], UMB-DB [127] are captured by Minolta Vivid 900/910 sensor. GavabDB [121] and FRAV3D [122] also use a laser digitizer called the Minolta VI-700 for capturing the databases. BJUT-3D [123] is captured by CyberWare 3030RGB/PS laser scanner which can simultaneously obtain both precise shape and colour texture. In addition, its illumination system reduces the variations resulting from the ambient light.

BU-3DFE [125] was captured by a 3dMD acquisition setup and combines both the passive and active stereo. The advantages of such systems are high accuracy and fast acquisition times. However, this setup is expensive and requires a time consuming calibration procedure and full collaboration of each individual. Texas 3DFRD [126] consists of 3D models that are captured using an MU-2 stereo imaging system (similar to the one used in BU-3DFE [125]) and contains 105 subjects.

The quality of 3D data captured by most sensors can be degraded by illumination variations and large variations can even result in the failure of the 3D shape estimation. Many manufacturers have devoted considerable effort to developing new technologies to solve this problem, for example by automatically adapting to illumination variations [27]. Artefacts might occur in range sensors even under ideal illumination conditions and the most common types of artefacts can be generally described as holes or spikes. 3D sensors have been improved in the recent years, which are less sensitive to ambient lighting, have fewer artefacts and require less explicit user cooperation [27].

Compared to other sensors that can capture the depth information with higher accuracy, the Photoface device uses the photometric stereo [9, 17]. It is a four light source photometric stereo acquisition system that can be located in the entrance of buildings, allowing relatively unconstrained captures. Each capture consists of four images with different light directions, from which surface normals and albedo can be calculated directly, with high resolution.

The use of albedo images can address the illumination problems found in many 3D data capture systems. Unlike other acquisition approaches, the Photoface device can estimate the 3D information with the minimal computational expense and its hardware is much cheaper than many other 3D capture devices, as it only requires one camera and four flash lights. The Photoface database therefore provides an ideal test source for face recognition algorithms designed for real-world applications [9, 17]. Nevertheless, the depth map of the Photoface captures is less accurate than that reconstructed from the 3D point clouds obtained by laser scanners, as it is reconstructed from surface normals. The regions with higher curvatures are flatter, which is hard to extract features for region-based recognition algorithms [18].

In addition, recently, there has been increasing interest in multi-modal 2D + 3D recognition algorithms [118, 119, 129-131] that seek to fuse both 2D (colour or texture) and 3D depth modalities in either the feature or decision level. These approaches typically demonstrate an improved recognition performance [27]. There is therefore a requirement for databases that provide suitable multi-modal data sets and this is met by most of the existing 3D face databases, which provide both 2D and 3D captures.

Table 3.2: Data acquisition methods of 3D face databases

Data Capture Method	**Databases**
Structured Light	3DRMA [119], York2 [124]
Inspeck Mega Capturor II 3D (Structured Light)	Bosphorus [13]
Minolta Vivid 700 (Laser)	GavabDB [121], FRAV3D [122]
Minolta Vivid 900/910 (Laser)	FRGC [12], ND-2006 [120], CASIA [97], UMB-DB [127]
CyberWare 3D scanner (Laser)	BJUT-3D [123]
Photometric Stereo	Photoface [9, 17]
3dMD (Stereo)	BU-3DFE [125]
MU-2 stereo imaging system	Texas 3DFRD [126]

Table 3.3: Widely used 3D data representations

Types of data	Characteristics
Range (Depth) Image	The z coordinates (distance from camera plane)of the face points are mapped on a regular x, y grid using linear interpolation
3D Mesh	3D object representation consisting of a collection of vertices and polygons
Point Cloud	A set of the 3D coordinates of the points of a face
Curvature	Each point in the face is described by its curvature (Gaussian, mean, principal)
Surface Normal	Each point of the face is described by its 3D (SN_x, SN_y, SN_z) unit normal vector

3.2.3 Database Population

The population of each database possesses some basic characteristics including the number of subjects and the range of captures per subject, the composition of subjects in terms of race, age and gender. As there are substantial differences in these characteristics over the databases under consideration, they can have significant influence on the performance evaluation. The size, age and race of the population of 13 databases are presented in Table 3.1. A database's size can be measured by the number of subjects and the number of captures. Figure 3.2 ranks 13 databases considered in order of their number of subjects. With the exception of the smallest (GavabDB [121]), these databases can be divided into two groups, medium-sized one with 100-143 subjects and large one with 350-880 subjects. ND-2006 [120] is clearly the largest in terms of subjects and also contains many captures for each individual. Information on the age range of the subjects is provided for just over half the databases in Table 3.1. Although the lower end of the age range is generally 18 and the upper end varies and is up to 75 for the Texas 3DFRD [126]. There are also some younger individuals such as a four year olds in the UMB-DB [127].

The ethnic composition of databases also has significant variations. Some databases mainly contain subjects from the dominant racial group of the country in which they were developed. For example, many Chinese individuals are included in CASIA [97] and BJUT-3D [123] while Caucasian individuals account for the majority in 3DRMA [119], GavabDB [121], FRAV3D [122], Bosphorus [13], UMB-DB [127] and Photoface [9, 17]. Databases with subjects from predominantly one race are not able to support performance evaluation across multi-ethnic groups. In comparison, FRGC v.2 [12], BU-3DFE [125] and Texas 3DFRD [126] contain individuals from many of the major ethnic groups and therefore provide the appropriate test data required to demonstrate the universally applicability of recognition algorithms.

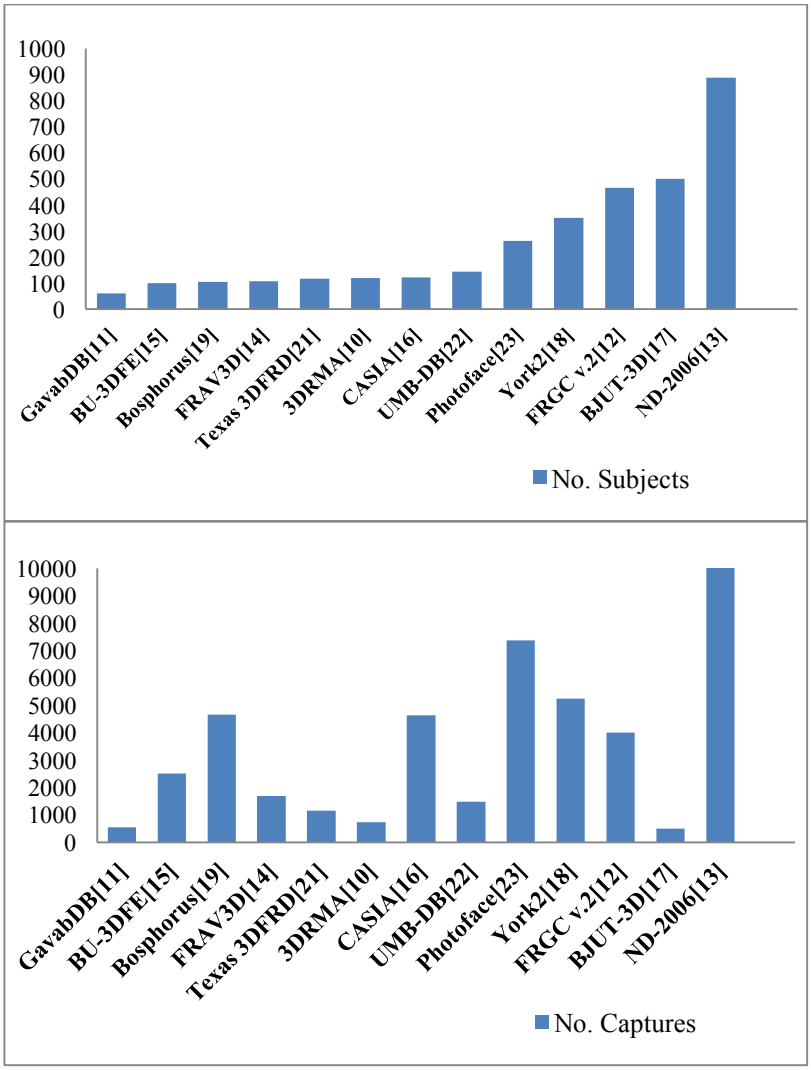


Figure 3.2: The number of subjects and captures for the databases.

Although many biometric methodologies have been shown to be capable of distinguishing between identical twins, the recognition of an individual twin using 3D facial data is still a challenging topic because twins often have virtually identical facial geometry. There are very few samples of twins in the 13 databases considered so far (e.g. only one pair in UMB-DB [127] and Photoface [9, 17]) and the inability of researchers to test their algorithms on identical twins has made it difficult to achieve further significant improvements in their recognition performance.

This limitation was overcome with the development of the 3D Twins Expression Challenge (3D TEC) database by the University of Notre Dame [132]. 3D TEC is the only database of 3D face scans in existence that has more than one pair of twins. It consists of 107 sets of identical twins (two of the triplets are included as the 107th set of twins) with captures of both a neutral and a smiling expression for each subject. The combination of factors related to the facial

similarity of identical twins and the variation in facial expressions make this a challenging 3D face database.

3.2.4 Pre-processing and Landmarks

Pre-processing is a vital step in 3D face recognition systems and its performance will significantly affect the recognition performance of the whole algorithm, for example the accuracy of the landmarking and feature extraction. The within-class similarity and between-class dissimilarity of some captures even might be lost without sufficient preprocessing work. The captures from some databases have been preprocessed to some extent or may provide manually detected landmarks, which can help researchers in feature space manipulation and landmarking performance evaluations.

Many face recognition algorithms require a set of landmarks located on the facial surface, from which relative geometrical relationships can be calculated [102, 133]. Landmarks are also often used to align the images and manually defined landmarks are also required to assess the accuracy of automatic landmark detection algorithms. Some, but by no means all, of the 3D face databases provide manually defined landmarks. The landmarks provided vary in both number and position and are therefore employed different application scenarios. Figure 3.3 presents the location of landmarks in the six databases from those in Table 3.1 for which a significant number of landmarks are provided.

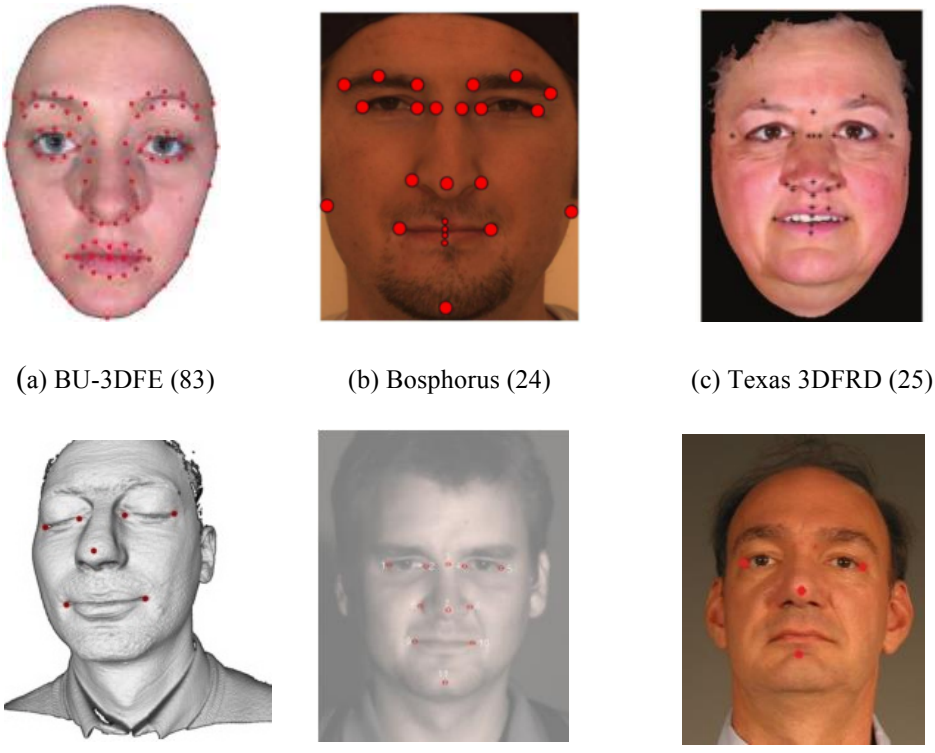
BU-3DFE [125] provides a set of metadata including the positions of 83 manually annotated feature points placed on each facial model, as depicted in Figure 3.3(a). These can be widely used in the development and evaluation of landmark-based face recognition algorithms. However, the location of tip on the nose region is not recorded in this database, which is arguably the most important landmark on the human face and is widely used in many face recognition algorithms. For example, the main face region can be cropped on the basis of the nose tip and it is also used to localize the nose region, which may contain sufficient information to recognize people. Since the nose tip plays a significant role in the face recognition, many methods are designed to detect it automatically. A database with accurate nose tip landmark will benefit this process.

The Bosphorus database [13] provides 24 manually labelled facial landmarks for each capture provided that they are visible in the given scan, see Figure 3.3(b). This includes all the significant landmarks, such as the nose tip and inner eye corners, which play a vital role in face

recognition algorithms. The Texas 3DFRD [126] provides a large number (25) of manually annotated anthropometric facial fiducial points for each capture, as shown in Figure 3.3(c). In addition, all its faces have been aligned to the frontal view and the nose tip is positioned at the image centre, which provide a good choice for 3D facial feature analysis. Compared to the Bosphorus database [13], many fiducial points around the nose region provided in the Texas 3DFRD [126] make it a useful database for nasal features extraction.

In the UMB-DB database [127], 7 landmarks including the nose tip, eye corners and mouth corners have been manually annotated, see Figure 3.3(d). Only the non-occluded feature landmarks are given for the captures with occlusions. Photoface [9, 17] provides two sets of metadata, created in terms of the x and y coordinates of 11 manually marked fiducial points, see Figure 3.3(e). Although UMB-DB and Photoface provide fewer landmarks than the Bosphorus and Texas 3DFRD, those they do provide are still useful for the alignment and landmarking evaluation.

Some of the other 3D face databases provide more limited landmarks information. For example, in FRGC v.2 [12] there are four landmarks as shown in Figure 3.3(f): two for both eyes, and one for the nose tip and the middle of the chin.



(a) BU-3DFE (83) (b) Bosphorus (24) (c) Texas 3DFRD (25)

(d) UMB-DB (7) (e) Photoface (11) (f) FRGC v.2 (4)

Figure 3.3: The landmarks provided by six databases with the number of landmarks shown in brackets

3.2.5 Modes of Variation

Databases that contain captures with variations in facial expressions, occlusions and poses provide more realistic and challenging test sets. The majority of 13 publically available databases, except the FRAV3D and BJUT-3D databases, contain variations of one type or another and these are detailed in Table 3.4.

3.2.5.1 Expressions

3D databases that contain substantial variations in the facial expression could significantly contribute to the performance improvement of 3D face recognition. This is the area of much current research interest and, in their excellent review paper, Smeets *et al*. provide a comprehensive overview of current 3D face databases with variations in expressions [111].

Table 3.4: Details of expression, occlusion and pose variations for 11 3D face databases.

Databases	Expressions	Occlusions	Poses
3DRMA [119]	Smile	Spectacles, beards and moustaches	Small up/down and left/right
FRGC v.2 [12]	Anger, happiness, sadness, surprise, disgust, puffy	Basic occlusions	Slight pose
ND-2006 [120]	Neutral, happiness, sadness, surprise, disgust and other	Basic occlusions	Slight pose
GavabDB [121]	Smile, frontal accentuated laugh, frontal random gesture	Hands and tongues	Left, right, up, down
BU-3DFE [125]	Anger, happiness, sadness, surprise, disgust, fear (in 4 levels)	No	Two views at approximately 45°
CASIA [97]	Smile, laugh, anger, surprise, closed eyes	Glasses	11 different types
York2 [124]	Neutral, happiness, anger, eyes closed, eye-brows raised	No	Uncontrolled up & down
Bosphorus [13]	35 expressions (action units & six emotions)	Beards, moustaches, hair, hands, and glasses	13 ground-truthed yaw and pitch rotations
Texas 3DFRD [126]	Neutral, smiling or talking faces with open/closed mouths and/or closed eyes	No	1
UMB-DB [127]	Neutral, smiling, angry and bored	Hair, glasses, hands in various configurations and positions, hats, scarves, and other miscellaneous objects	1
Photoface [9, 17]	Smile, surprise, open mouth, scream and others	Glasses, facial hair (beards, moustaches and stubble)	1 frontal - 5 profile

The most popular 3D face database is FRGC (cited by 1846 articles), which is also a very challenging database [12]. Compared with the training set, the validation set contains plenty of expressions, such as anger, happiness, sadness, surprise, and disgust, as well as puffy faces. ND-2006 contains a greater range of expressions than FRGC v.2, in terms of six different types of expressions [happiness (2230), sadness (205), surprise (180), disgust (389) and other (557)] in addition to 9889 neutral expressions [120]. Even though the neutral category contains the largest number of images, expressions account for more than a quarter of the database.

Although ND-2006 offers multiple expressions, BU-3DFE was arguably the first dedicated expression database available for the research community [125]. BU-3DFE has four levels (low, middle, high and highest) of 6 expressions captured in two views at approximately 45°. Including the neutral expression, this gives 25 3D expressions for each subject and results in a database of 2,500 3D facial expressions. BU-3DFE therefore enables researchers to evaluate the degradation of recognition performance with increasing strength of expressions and has contributed to an increased understanding of facial behaviour and the fine 3D structure inherent in human facial expressions.

The GavabDB database contains 2 frontal and 4 rotated images with neutral expressions and 3 frontal images in which the subject presents different facial expressions (smile, laugh and a random gesture chosen by the individual), giving a total of 9 captures per subject [121]. Compared to GavabDB, the York2 database provides more captures per subject with variations, including 15 captures with diverse expressions (neutral, happiness, anger, eyes closed, eye-brows raised) and poses [124]. The variations in expressions and poses of both databases occur independently.

The CASIA database also considers variations in the illumination. In addition to individual variations of expressions, poses and illumination, each subject also contains combined variations of expressions (smile, laugh, anger, surprise, and closed eyes) under different lighting conditions and poses [97]. The Photoface database displays variations of expressions (> 600 smiles and > 200 surprises, open mouth, screams) [9, 17].

The Bosphorus database provides a rich set of expressions, systematic variations of poses and different types of occlusions. The facial expressions are composed from a selected subset of 35 Action Units as well as six basic emotions. Some professional actors are used as they are considered to provide more realistic expressions [13]. Compared to the FRGC database, captures in the Bosphorus database are less noisy and have more intense expression variations.

3.2.5.2 Poses

Many 3D face databases contain captures with different kinds of pose variations. Compared with others, the Bosphorus database possesses three types of variation in pose corresponding to seven yaw angles (+10°, +20°, +30°, +45°, +90°, -45° and -90°), four pitch angles (upwards, slight upwards, slight downwards, downwards) and two cross-rotations (right-downwards and right-upwards) which incorporate both the yaw and pitch [13]. Furthermore, accurate angles are provided for each pose and in comparison to other 3D databases with different poses, these angles provide ground-truthed information for research.

Some other 3D face databases contain poses where the approximate directions are provided. GavabDB [121] contains 4 rotated images (right and left profiles at ~90º, and looking up and down at ~35º) without expressions. BU-3DFE [125] has four levels of 6 expressions captured in two views at 45°. CASIA [97] provides 11 different poses under certain range of directions, e.g. left 20º-30º. York2 [124] includes 6 different captures per subject with poses (45° in all directions, 90° to the left and right). The Photoface database [9, 17] contains poses from 1-5 where 1 is frontal and 5 is extreme pitch, roll or yaw.

3.2.5.3 Occlusions

There are many unconstrained and challenging scenarios where faces are partially occluded by various objects such as hair, glasses, beards, moustaches, hands and tongues. It is only relatively recently that these occlusions have begun to be considered by 3D face recognition techniques but, as the interest in this area increases, so will the demand for suitable databases. Table 3.4 shows that several 3D face databases contain captures with various types of occlusions.

The Bosphorus 3D face database provides 381 occluded faces from 105 subjects with three types of occluded objects (hands, glasses and hair). Different parts of the face are also occluded such as the mouth, one eye (by hands), both eyes (by glasses), and part of the forehead or face (by hair) [13].

UMB-DB is particularly designed to evaluate facial occlusions and, in comparison with the Bosphorus database [22], it has a larger number of occluded faces. UMB-DB provides an increased variability in terms of the location of the occluded regions with an average extent of occlusion of 42%, up to a maximum of 84%. In total, there are 590 captures with occlusions from various real-world objects, see Table 3.4. Other databases that contain occlusions include

3DRMA [119] (spectacles, beards and moustaches), Photoface [9, 17] (glasses, facial hair) and GavabDB [121] (hand and tongue).

In addition to occlusions caused by objects, self-occlusions are also found in captures that have large head rotations, particularly in the yaw direction. Therefore, self-occlusions are found in those databases with wide variations in pose that were discussed in the previous section.

3.2.6 Baseline Performance Results for the FRGC and Photoface Databases

"Baseline performance serves to demonstrate that a challenge problem can be executed, to provide a minimum level of performance, and to provide a set of controls for detailed studies" [12]. Some methods that can be selected as baseline algorithms for face recognition include PCA, LDA or FDA, ICA and KFA. Therefore, the baseline performance reported on publically available databases can give researchers a good reference to evaluate their algorithms. In this chapter, a brief summary of two 3D databases will be presented, one from each of two categories of databases: true and reconstructed 3D captures.

FRGC v.2 provides both large 2D and 3D training dataset. The still images training set consists of 12,776 images from 222 subjects, with 6,388 controlled still images and 6,388 uncontrolled still images, while the 3D training set contains 3D scans, and controlled and uncontrolled still images from 943 subject sessions. In version 2.0, six experiments are designed to demonstrate recognition performance by the baseline evaluation, with emphasis on 3D and high resolution still imagery. Experiments 3, 5 and 6 examine different potential implementations of 3D face recognition. Experiment 3 measures the recognition performance when both the enrolled and query images are 3D and Experiment 5 and 6 explore one potential scenario for 3D face recognition, where the enrolled images are 3D and the target images are controlled or uncontrolled still 2D images.

Table 3.5: Baseline performance reported in FRGC and Photoface papers

Databases	**Method**	**Training and testing samples**
FRGC v.2 [12]	Six PCA based experiments (3, 5 and 6 are related to 3D face recognition)	Large still and 3D training captures from 222 and 466 subjects
Photoface [9, 17]	PCA and NMFC based face recognition using albedo, depth and normalface images (albedo and depth images can be obtained by different methods)	**One Image**: training (126 probes with one capture) and testing (126 clients and 135 impostors with one capture) **Two Images**: training (96 probes with two captures) and testing (96 clients and 135 impostors with one capture)

Compared to other databases, the aim of baseline experiments in the Photoface database is to assess the recognition performance of all three modalities (i.e., albedo image, normals and depth maps) using fusing strategies [9, 17]. Experiments can be divided into two different parts and they evaluate the performance of "One sample face recognition", named "One Image" which is the most challenging face recognition scenarios [134]. For "One Image", the training set (or client set) was defined by the 126 subjects and the remaining 135 subjects with one image per person are considered to be impostors. This is a very challenging experimental procedure and exploits only the grayscale albedo image, the surface normals derived from the application of photometric stereo, and the depth image derived from the integration of the normal field. Photoface also takes Two Images experiments (two images from 96 subjects for training while the third is used for testing client claims) into consideration for making a comparison with One Image experiments. Zafeiriou *et al*. demonstrated that multimodal decision fusion, performed by combining the matching scores for each person across the modalities of 2D, albedo and depth image, was found to produce the best performance in the Photoface database [9, 17].

With increasingly more accurate and effective 3D face reconstruction methods developed in recent years, it brings more potentials to design the face recognition algorithms for reconstructed 3D captures. A good example is photometric stereo, the method used in the Photoface database. As mentioned in Section 3.2.2, this device meets the commercial requirement and can be easily used in real-world application scenarios. However, compared with other databases that contain higher resolution 3D captures, feature detection and extraction on this kind of captures are still very challenging, which require more novel and complicated approaches to improve the recognition performance rather than reuse the existing strategies for high resolution captures.

3.3 3D Face Databases Used in this Thesis

The FRGC database contains a large number of captures with many variations in the expression and pose and some basic occlusions. It is considered by some researchers to be the most challenging database for face recognition [111]. For Experiment 3 in [12], the 3D captures in the Spring2003 folder, consisting of 943 captures from 275 subjects, are used for training and the other two folders (Fall2003 and Spring2004) consisting of 4007 captures from 466 subjects with more variations are used for testing. The combination of these two folders is known as the FRGC v.2. To extract more effective discriminative features for expression robust face

recognition, the FRGC v.2 captures are manually divided into neutral and non-neutral parts. Those expressions that only cause minor changes to the face surface are considered as neutral and the resulting division produced neutral and non-neutral sets of size 3396 and 1554, respectively. Although nearly all the captures in Spring2003 are neutral, the training set is still defined as "All". The testing set is either considered as "All" or divided into "Neutral" and "Non-neutral" and these three scenarios, "All vs All", "All vs Neutral" and "All vs Non-neutral", are used in the following chapters.

The Bosphorus database contains many kinds of action units from the Facial Action Coding System, which can represent different changes on the face surface caused by human muscle movement, and also includes the basic universal expressions. Nearly all the captures contain different degrees or types of expressions and thus evenly dividing them into the training and testing sets guarantees that the various expressions are evenly distributed in both subsets. This scenario is termed "All vs All" and there are 2913 captures in total. As this work aims to find the most discriminative and expression robust features, only those captures without large pose variations or occlusions are considered.

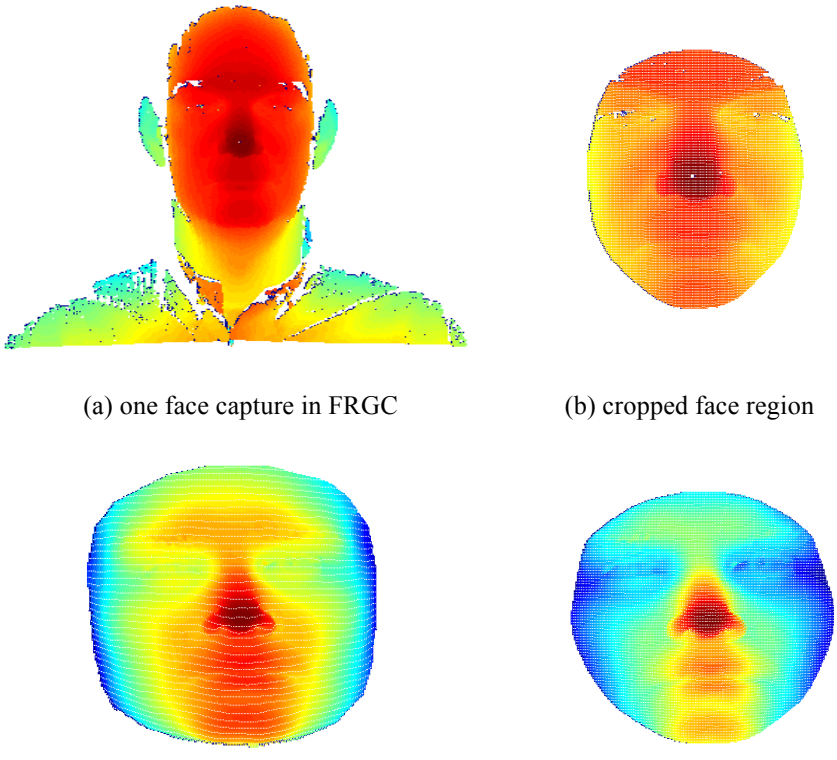
The Photoface database consists of 3187 captures of 453 subjects for the final version and only 1839 captures of 261 subjects are used in [17] and [9]. For the comparison of the identification and verification results, the same database population as [9, 17] is used. The Photoface capture device was located in a workplace corridor and employees walked through it, which requires a lower degree of cooperation in comparison with other 3D databases [18]. The Photoface database provides researchers with the opportunity to test the recognition performance on such naturally captured data and most captures contain expression, pose and occlusion variations.

An extension of the Photoface database, the 3D Expression-VISible and InfraRed (3DE-VISIR) database, is proposed in [18], which contains 363 relatively unconstrained captures of 115 subjects. Compared to the Photoface data captured by the visible light, each capture in the 3DE-VISIR database considers one session captured by the near infrared light and another session by the visible light. In particular, the aim of this new database is evaluating the recognition performance under expressions, so that for each subject there are at least three captures of different expressions (positive, negative and neutral).

3.4 Preprocessing of the FRGC and Bosphorus Databases

Due to the limitations and variations of existing 3D data capture devices, many captures contain artefacts such as spikes, holes and missing data, degradations that are particularly prevalent in the FRGC Spring 2003 folder. These artefacts significantly affect the subsequent processing and degrade the final recognition performance [135]. As a consequence, this work employs a three-step preprocessing scheme. First, the nose tip is detected by finding the largest convex region on the face and then the main face region is found by intersecting its surface with a sphere centred on the nose tip. Then, three typical filters are applied sequentially for denoising, namely the spike removal filter [5], hole-filling filter and smoothing filter. In addition, nearly all the captures within the databases contain small or large pose variations which can degrade the within-class similarity of each subject. To address this problem, an improved PCA based alignment algorithm is used to correct the pose variations [5, 7]. These stops are described below.

3.4.1 Face Cropping



(a) one face capture in FRGC (b) cropped face region

(c) one face capture in Bosphorus (d) cropped face region

Figure 3.4: An illustration of 3D face cropping: (a) and (c) show the original 3D captures in the FRGC and Bosphorus databases, respectively; (b) and (d) are their corresponding face cropping results

The captures in most 3D face databases are not always cropped correctly and invariably contain some irrelevant details as shown in Figure 3.4(a) and (c). Consequently, to further refine the detected or cropped facial region, the nose tip is first detected and then the face region as shown in Figure 3.4(b) and (d) is found by intersecting a sphere with radius 80 mm centred on the nose tip [5]. The nose tip is detected by finding the largest convex region on the face, which can be found by thresholding the SI using $-1 < SI < -5/8$ to produce a binary image [7]. The largest connected component is detected and the centroid is saved as the tip.

3.4.2 Denoising

Compared to 3D data provided in the Bosphorus and Photoface datasets, the captures in FRGC have more artefacts represented as spikes, holes and missing data, as shown in Figure 2.4(a) and Figure 3.4(a). These artefacts might significantly affect the subsequent recognition steps and degrade the final recognition performance. As a consequence, to remove the noises in FRGC, cubic interpolation is first used to correct the missing data in x, y and z coordinates and then spike removal, hole-filling and smoothing filters are sequentially used in the denoising phase.

3D face captures are usually degraded by the impulsive noise, holes and missing data, which are more salient on the Z map. The outlier points are first found by standard deviation calculation in the Z map that uses a 3×3 sliding window [5]. Those distances that are greater than a threshold d_t from any one of its 8-connected neighbours. d_t is automatically calculated using $d_t = \mu+0.6\sigma$ (where μ is the mean distance between neighbouring points and σ is its standard deviation). The same interpolation procedure is employed to fill the holes again in the Z map. Then, the morphological filling is applied to the depth map and a larger threshold is selected to preserve the natural holes on the face, in particular near the eye corners. Finally, median filtering is used to remove the impulsive noise on the face surface. Figure 2.4(b) shows the denoised capture of (a).

3.4.3 PCA Based Alignment

Nearly all the captures in the FRGC database contain small or large pose variations which can significantly affect the final recognition performance. For the Bosphorus database, some frontal viewed captures still contain slight pose variations, even though all the captures are labelled with the ground-truthed rotations. To address this problem, the PCA based alignment method

proposed by Mian *et al*. [5] is used to correct the pose of the captures. This approach use a $3\times n$ matrix () of the *x*, *y*, and *z* coordinates of the 3D face data to generate *P*.

$$P = \begin{bmatrix} x_1, x_2 \dots x_n \\ y_1, y_2 \dots y_n \\ z_1, z_2 \dots z_n \end{bmatrix}$$

The points' mean (*m*) is translated to the origin and their 3×3 covariance matrix (*C*) is calculated.

$$m = \frac{1}{n}\sum_{k=1}^{n} P_k$$

$$C = \frac{1}{n}\sum_{k=1}^{n} P_k {P_k}^T - mm^T$$

The points are then mapped onto the principal axes after multiplying them by the covariance matrix's eigenvectors: $P' = C(P - m)$. They are then uniformly resampled with a resolution of 0.5 mm. The missing points caused by self-occlusion, especially around the nose region, are replaced by the 2D cubic interpolation. The procedure processes iteratively until *C* converges to an identity matrix and the nose tip is also re-detected. Furthermore, the face will be cropped again on the new 3D data. After the alignment procedure is completed, a small constant angular rotation along the pitch direction is added to the face pose as this helps the landmarking algorithm to detect the nose region.

3.5 Preprocessing of the Photoface Database

Denoising is always considered as a significant step for preprocessing the data for the reason that captures from 3D devices contain some holes, spikes and missing parts to some extent which will significantly affect the following preprocessing steps and the consistency of the feature extraction. Those kinds of noisy data are prominent in the FRGC database and less important for the Photoface database. Therefore, it is not necessary to employ all the denoising steps for the Photoface, as all 3D captures in this database are reconstructed from their corresponding surface normals and have been smoothed during the reconstruction process.

In a similar manner to FRGC and Bosphorus databases, the nose tip is automatically detected by thresholding the SI and the main facial region is cropped by intersecting a sphere centred on the detected nose tip. For 3D captures in the FRGC and Bosphorus databases, captures can be estimated with high resolution and many facial local details can be preserved. However, in

the Photoface database the depth information is reconstructed from the surface normals and is distorted to some degree. The resulting 3D captures look somewhat flat compared to real 3D model. Moreover, as the resolution maps are not available, the face cropping is more complicated, which requires methods that utilise prior information on the facial shape. Given the same resample setting as real 3D captures, a radius of 80 mm, which is widely used in the FRGC and Bosphorus databases, is too small for the Photoface captures. 150 mm has been found to be an appropriate setting by observing that can approximately crop the main face region. It is important to note that the radius used is not critical to the recognition performance as the subsequent features are only extracted from the nasal and adjoining regions.

The nose region is cropped by finding the intersections of three cylinders, each centred on the nose tip [7]. Similarly, the coefficients of cylinders used in the Photoface database are different from FRGC and Bosphorus. In [7], radii of 40 mm and 70 mm are suggested for horizontal cylinders and 50 mm for vertical cylinder for the FRGC and Bosphorus database. For the Photoface database, they are set to 70 mm, 120 mm, 100mm, 100 mm, respectively, which provides the most consistent results. Finally, according to the location of nose tip, an improved PCA based alignment algorithm is used to address the pose variations on both the nose and cropped face region [5, 7]. During each iteration, the nose tip is relocated.

3.6 Baseline Recognition Performance Evaluation

To provide a comparable performance contrast among the FRGC, Bosphorus and Photoface databases, a subset of 18 subjects, each with 20 captures, is selected from each database. The captures for each subject are then split into 10 captures for training and 10 captures for testing. For the FRGC database, all the 360 captures are selected from FRGC v.2, Fall2003 and Spring2004 folders, as these two folders contain plenty of expression, pose and occlusion variations. For the Bosphorus database, to provide a direct comparison with features extracted from FRGC, only those captures without large pose variations or occlusions are considered, and these captures contain plenty of expressions. Similarly, nearly all the captures contain natural variations in expressions, poses or occlusions (glasses and scarves) in the Photoface database.

The baseline performance using PCA and LDA is evaluated under both identification and verification scenarios. Also, different kinds of 3D data representations [depth, surface normals, SI and principal_curvature_minimum (k_{min})] are explored in this evaluation. The R_1RRs of identification and EERs of verification scenarios are illustrated in Table 3.6.

Table 3.6: Baseline identification and verification performance evaluation on the subset (18 subjects, each with 20 captures) of the FRGC, Bosphorus and Photoface databases. The features are extracted from depth, surface normals (SN_x, SN_y *and* SN_z), SI and k_{min}, respectively.

	Algorithm	Scenarios	Depth	Surface Normals			SI	k_{min}
				SN_x	SN_y	SN_z		
FRGC	PCA	R_1RR	72.07%	66.48%	73.74%	79.33%	70.95%	73.18%
		EER	20.13%	28.48%	14.53%	21.72%	22.35%	16.74%
	LDA	R_1RR	96.65%	94.41%	97.21%	98.32%	93.85%	91.62%
		EER	1.68%	2.79%	1.68%	1.12%	2.23%	4.35%
Bosphorus	PCA	R_1RR	64.25%	41.90%	73.74%	62.57%	44.69%	54.24%
		EER	25.71%	35.70%	18.44%	24.60%	37.41%	31.48%
	LDA	R_1RR	87.71%	83.80%	91.62%	89.94%	69.83%	81.36%
		EER	7.92%	6.65%	3.91%	6.70%	11.76%	6.78%
Photoface	PCA	R_1RR	55.29%	58.82%	83.53%	64.12%	71.18%	76.47%
		EER	22.94%	25.80%	14.12%	20.00%	21.80%	18.82%
	LDA	R_1RR	94.71%	94.12%	96.47%	91.18%	92.35%	95.29%
		EER	2.35%	2.94%	1.76%	4.12%	2.82%	2.35%

Instead of using the whole human face, one region-based algorithm proposed by Emambakhsh *et al*. [7] is employed for feature extraction: All the original captures are pre-processed sequentially by nose tip detection, face cropping, denoising and alignment. To build the feature space, 16 landmarks on the nasal region are automatically detected and 75 curves found by connecting selected two landmarks. This provides a simple and effective representation of the nasal region.

For different types of data, the three databases produce diverse performances, which is mainly due to the properties of data itself and the facial variations discrepancy. Specifically, when using the PCA classifier, features extracted from the depth information from the FRGC and Bosphorus databases are clearly more discriminative than the Photoface database, for the reason that some distortions may occur during reconstruction phase. However, the SN_y calculated from the reconstructed depth map outperforms the other two datasets. In addition, SI and k_{min} information calculated from Bosphorus contribute less to recognition performance.

3.7 Conclusion

This chapter has reviewed and evaluated a selection of the publicly available 3D face databases that have been developed over the past decade. These databases underpin many of the recent advances in face recognition, for example in robustness to expressions and occlusions, by providing researchers with meaningful databases on which to evaluate their algorithms. A brief introduction to 13 of the most widely used publically available 3D face databases has been provided. Some of the factors which need to be considered when attempting to improve the recognition performance, in particular to real-world scenarios, are also discussed. By considering the database characteristics of Data Capture, Database Population, Pre-processing and Landmarks and Modes of Variation, a database classification methodology has been proposed. The classification methodology has been used to categorise the databases previously introduced, providing researchers with a useful resource to help choose appropriate databases on which to evaluate their recognition algorithms.

According to varying characteristics of 3D face databases, FRGC, Bosphorus and Photoface are selected to be used in the following chapters. Two different preprocessing methods are implemented on the two true 3D (FRGC and Bosphorus) and reconstructed 3D (Photoface) databases, respectively. Baseline recognition performance results are then presented in Section 3.6 show that the captures from different databases possess varying types of 3D representations, which might subsequently produce varying recognition performance.

Chapter 4 Matching Curves for 3D Face Recognition

4.1 Introduction

Facial expression is an intrinsic source of variability in human faces. Expressions deform the face surface to a greater or less extent, according to their type or degree of expressions. In the past decade, expression variations have been one of the most challenging sources of variability in 3D face recognition, especially for scenarios where there are a large number of face samples to discriminate between. Some relatively stable structures and patches on the face can be used to help design expression invariant face recognition algorithms. A good example is the nose region which, when compared to the forehead, mouth and eyes, is more consistent over different expressions. The nose is also very difficult to be deliberately occluded by hair, hands and scarves [112]. In addition, using convexity it is relatively straightforward to detect and segment the central region of the 3D face that contains the nose [7]. Therefore, feature extraction on the nose region offers many advantages for expression robust 3D face recognition.

Although many methods can be used for extracting features from the 3D face, depth information is commonly used because of its efficiency. One of the most popular approaches is to draw curves between extracted landmarks. Drira *et al*. [49] and Ballihi *et al*. [80] achieved very good recognition performances using various sets of facial curves. However, the location of the curves covered some expression sensitive regions, presenting problems for the expression robust recognition. In addition, the nose tip had to be detected manually. Emambakhsh *et al*. automatically detected four significant landmarks (the nose tip, the saddle and the two alar) and used these landmarks to define a set of 28 nasal curves that were used for expression robust recognition [7]. This work only considered landmarks and curves directly on the nasal surface and didn't include the regions adjoining the nose, between the cheek bones and nasal bridge. Indeed, these regions are also relatively more stable and less affected by occlusions when compared to other patches on the face surface and contain additional discriminative features to those on the nose. In this chapter, the nose and its environs are defined as the "cheek/nose" region, from which features are selected to form the feature space.

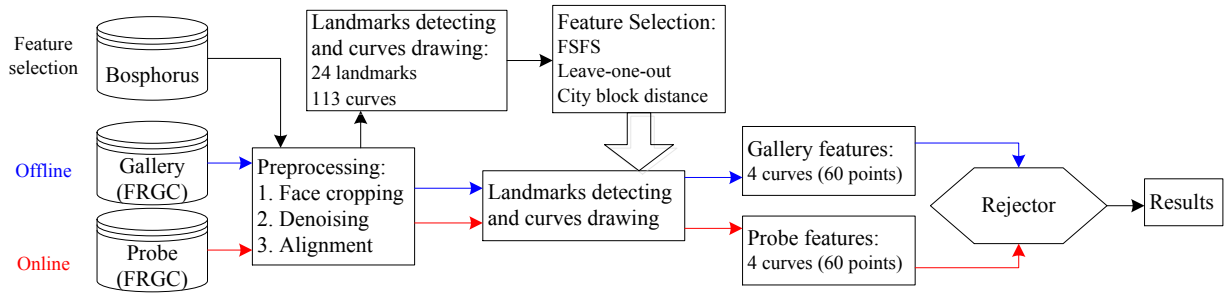


Figure 4.1: The block diagram of the proposed rejector.

In addition to overcoming the problems caused by expression variations, another challenge for 3D face recognition is the ability to discriminate between a large number of classes. Pattern rejection, proposed by Baker and Nayar [136], is an efficient and effective approach to improve the classification performance, particularly for this scenario. Mian *et al*. used a 3D SFR combined with a SIFT descriptor to form a rejector whose combined classification performance is very high [5]. However, it required both 2D and 3D features and the rejector's performance is much reduced when only the 3D SFR features are considered.

In this chapter, a new expression robust 3D face recognition algorithm is proposed by matching curves on the nasal and adjoining cheek regions. After pre-processing, 24 landmarks are localized on the cheek/nose region and a set of 113 curves joining the landmarks are defined. Feature selection using the Bosphorus database identifies just 4 curves, and further experiments determine that each curve only requires 15 points, to produce a rejector that quickly and effectively eliminates a large number of ineligible candidate faces from the gallery. A block diagram of this rejector is shown in Figure 4.1. In addition, the recognition performance of the radial and root curves extracted on the nasal and adjoining regions are investigated in this chapter.

4.2 Cheek/Nose Region Landmarking and Curves Finding

For region based face recognition algorithms, the accuracy of landmark localization plays a significant role in the feature extraction, which will directly affect the within-class similarity and between-class scatter. However, high accuracy of landmarking is always accompanied by more complicated algorithms or more computation time, making a difficult compromise for real time automatic face recognition systems. The landmarking strategy proposed in [7] provides a good solution to address this problem.

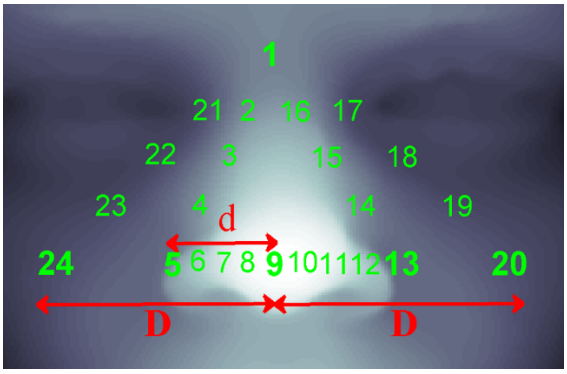


Figure 4.2: 24 automatically detected landmarks on the cheek/nose region. L9, L1, L5, L13, L20 and L24 are found and the reminder are found using proportions.

As shown in Figure 4.2, the 16 landmarks proposed in [7] are evenly distributed on the boundary of the nasal region. To further investigate the adjoining region of the nose and cheek bone, 8 more landmarks are localized. Of these, L9, L1, L5, L13, L20 and L24 are detected directly and the remainder are found by proportional division of the lines joining the landmarks previously found. The method proposed in [7] is first used to locate four main landmarks L9, L1, L5 and L13. The approaches to find these are described below.

4.2.1 Tip (L9)

In the preprocessing step, the nose tip has been approximately localized by finding the largest convex regions on the thresholded SI map (SI_{map}). To avoid the spike noise which is not removed in the denoising phase, the morphological dilation and median filter are used for further smoothing the candidate tip region. Specifically, the original thresholded SI map (SI_{map_o}) is dilated with a disk structuring element (B) (radius is 3) and multiplied by the depth map (Z_{map}). The spike noise on the resulting Z_{map} is removed by median filter and the maximum point on the Z_{map} is considered as the final nose tip.

$$\begin{cases} SI_{map} = SI_{map\_o} \oplus B \\ Z_{map} = Z * SI_{map} \end{cases} \quad (1)$$

4.2.2 Root (L1)

Once the location of the nose tip is fixed, the candidate root points can be found by using a set of planes which are perpendicular to the x-y plane. By intersecting the planes with the nose surface, many curves passing through the nose tip are obtained. The angles between the planes and y-axis are set from -15˚ to 15˚. The candidate nasal root points are located at the bottom of each curve and the final root is found at the maximum of those minima.

4.2.3 Alar Grooves (L5 and L13)

Similar to the method used for the nasal root detection, the candidate alar grooves are first located by finding curves passing through the nose tip. With the angular deviation [-15˚, 15˚] to the *x*-axis, a set of planes which includes the nose tip and are perpendicular to the *x*-*y* plane are intersected with the nose surface. On the basis of the special characteristics of the nasal shape, candidate alar grooves can be located at the lower part on each curve. In particular, the depth value of candidate points shows a sudden drop in comparison with their neighbouring points.

Therefore, to find alar grooves, each curve is differentiated and the candidate landmarks can be found by the maximum value of each curve's gradient. However, as all the points on the curves are used for gradient calculation this method is very sensitive to the noisy data and it is also affected by the location of nostrils. To address this problem, K-means clustering is used to divide the candidate alar grooves points and outliers into two groups, using the distance from candidate points to the nose tip.

4.2.4 Cheek Landmarks (L20 and L24)

The cheek landmarks used in this chapter define the extremities of the cheek/nose region. Two methods are employed to accurately and consistently locate them on the facial surface and both methods extend the line between the nasal alar grooves and the nose tip horizontally. The first approach sets the distance *D* in Figure 4.2 as a "*Proportional distance*" of L5 to L9 [*d* in Figure 4.2]. With this approach, the length from L9 to L24 depends on the size of the subjects' nose, such that each capture in the database potentially has different proportional distances. The second method directly sets L20 and L24 using a "*Constant distance*" from L9 in the horizontal direction, for example 40 pixels.

4.2.5 The Remaining Landmarks and Curves Drawing

The remaining landmarks are found by dividing the lines (L1 to L5, L5 to L9, L9 to L13, L13 to L1, L1 to L20 and L1 to L24) into four equal segments and the positions of the resulting points are shown in Figure 4.2. Different numbers of segments were investigated and it was found that this was not a major contributory factor in the overall recognition performance, with four segments offering a good compromise between the complexity and accurate representation of the cheek/nose region.

4.3 The Definition of the Pattern Rejector

The effectiveness and efficiency of recognition systems are the crucial challenges of recognition algorithms, especially for the increasing number of samples enrolled. Baker and Nayar proposed the notion of pattern rejection, to greatly enhance the efficiency in two aspects of a recognition system: the case with a large number of classes to discriminate between and when the recognition must be performed a large number of times [5]. It is generalized from the basic classifier which can quickly eliminates a large fraction of the candidate classes or inputs.

Pattern rejector is defined as "an algorithm ψ that given an input, $x \in S$, returns a set of class labels $\psi(x)$", where x is a measurement vector and S is the classification space. The effectiveness $Eff(\psi)$ of a rejector is the expected cardinality of the output $E_{x \in S}(|\psi(x)|)$ divided by the total number of classes, M [136]. The value of $Eff(\psi)$ is inversely proportional to the classification performance and, for face recognition, M is the size of gallery.

$$Eff(\psi) = \frac{E_{x \in S}(|\psi(x)|)}{M} \quad (1)$$

Compared to considering all the possibilities in the recognition system, this new approach allows the algorithms to dedicate their efforts to a much smaller subset. Also, the rejectors designed by different algorithms can be combined to build a composite rejector, which will further boost the recognition performance.

4.4 Curves Drawing

Using the landmarks shown in Figure 4.2, the intersection of planes passing through any two pairs of landmarks on the facial surface results in a total of 113 curves as shown in Figure 4.3. To normalize all the curves, all the original 3D data are translated to the nose tip and the number of the points of each curve is initially resampled to 50 points.

The 75 nasal curves are found by connecting the following landmark pairs in Figure 4.2: L1 to L2-L8 and L10-L16; L2 to L6-L8 and L10-L16; L3 to L16, L10-L15 and L6-L8; L4 to L14-L16, L10-L13 and L6-L8; L5 to L13, and L6-L7; L9 to L1-L5 and L13-L16; L14 to L5-L8, L10-L12; L15 to L5-L8, L10-L12; L16 to L5-L8, L10-L12. Also, the 38 cheek/nose curves are found by connecting: L17-L20 to L13-L16; L21-L24 to L2-L5; L1 to L20 and L1 to L24; L17 to L21, L18 to L22, L19 to L23 and L20 to L24.

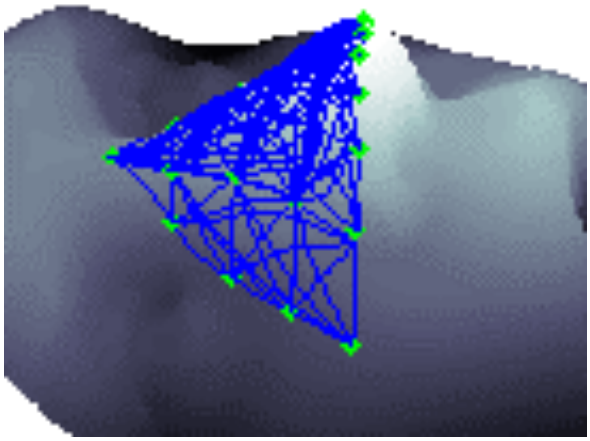


Figure 4.3: 113 curves found on the cheek/nose region

4.5 Feature Selection Using the Bosphorus Database

4.5.1 Motivation and Feature Selection

Instead of using the whole face region, extracting features on the cheek/nose region, which is relatively robust under various expressions, provides an effective approach for 3D face recognition. The 113 curves found in Figure 4.3 provide a comprehensive coverage of the cheek/nose region. However, selecting all the curves as features produces a high dimensionality feature vector which is unlikely to have a good classification performance [95, 96]. This so-called "curse of dimensionality" requires a large number of training samples to be added to the classifier when the dimensionality increases. To address these, feature selection is used to find a subset of curves that can produce a high recognition performance.

There are a lot of feature selection methods for dimensionality reduction. FSFS is applied to reduce the feature space dimensionality in this chapter for the reason that it provides a simple and effective way of selecting the curves step by step. FSFS starts from an empty set and first selects the best single curve, in terms of classification performance. Then, other curves which produce the highest R_1RR based on the leave-one-out method, when combined with the previously selected curves, are iteratively added to the set. The recognition rate of each single curve is calculated by the nearest neighbour city-block distance that has been shown to have a better discriminatory power when the feature space is sparse [74]. The feature selection employs the Bosphorus database, as the majority of its captures contain little noise, enabling relatively accurate landmarking and curve drawing.

4.5.2 Feature Selection Results of 75 Nasal Curves

The feature selection results of 75 curves are shown in Figure 4.4 and the number of points on each curve is resampled to 50. To produce the best recognition performance, 21 curves are selected by FSFS to build a feature vector whose R_1RR reaches to 82.5%. This implementation outperforms the result in [7] to some extent as they achieved the highest R_1RR, ~82.5%, when 28 curves are selected, which are using more curves than this implementation. Furthermore, for the recognition performance of each single curve, the most discriminative curve is the nasal bridge one (L9L1) with the highest R_1RR, 45.65%, which is much higher than the one recorded in [7], ~38%.

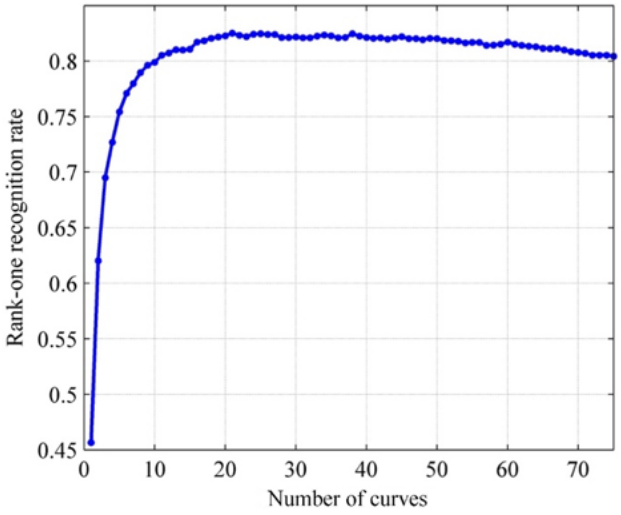


Figure 4.4: R_1RRs against the number of curves selected by the FSFS algorithm using 75 nasal curves

4.5.3 Feature Selection Results of 38 Curves on the Cheek/Nose Region

Using the detected cheek landmarks and their neighbouring nasal landmarks, 38 curves are found mainly on the cheek region for recognition performance evaluation. A constant distance of 40 pixels, was used to localize two cheek landmarks, which is a reasonable setting for 3D captures in the Bosphorus database and could be further evaluated in the following sections. R_1RRs against the number of curves selected by the FSFS algorithms is shown in Figure 4.5 and Table 4.1. To achieve the highest R_1RR, 6 curves which produce 82.77% of the R_1RR are selected by the FSFS method. It is higher than that of 21 nasal curves (82.5%), which produce the highest R_1RR in 75 nasal curves selection. If only using 4 curves, they achieve 82.5% of R_1RR, which is the same as the performance of 21 nasal curves but using a very small sized feature vector. This result is very promising and shows great potential of extracting expression invariant features on the adjoining regions between the nose and cheeks.

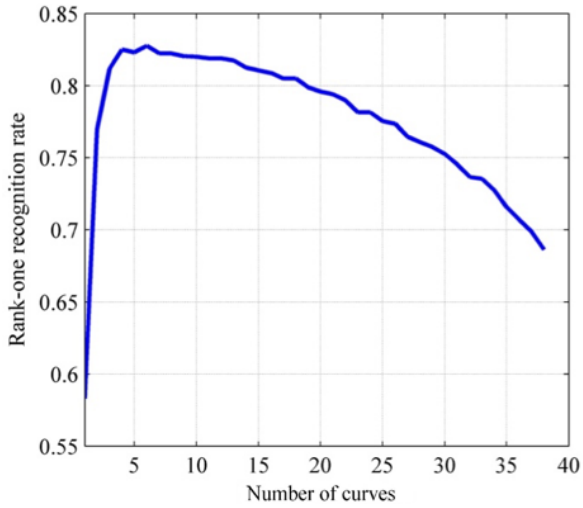
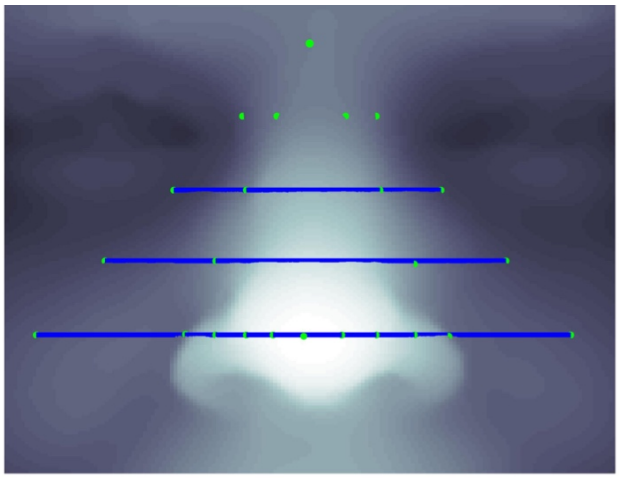
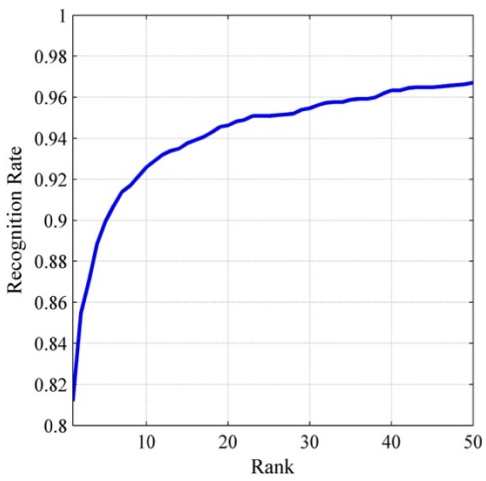


Figure 4.5: R_1RRs against the number of curves selected by the FSFS algorithm using 38 curves

Table 4.1: R_1RRs using the curves selected by FSFS

	Curves selected by FSFS	**R_1RR**
1	L18L22	58.31%
2	L18L22, L20L24	76.98%
3	L18L22, L20L24, L19L23	81.18%
4	L18L22, L20L24, L19L23, L1L20	82.50%
5	L18L22, L20L24, L19L23, L1L20, L17L21	82.31%
6	L18L22, L20L24, L19L23, L1L20, L17L21, L16L18	82.77%





(a) Three horizontal curves (b) CMC curves

Figure 4.6: Three horizontal curves across the cheek/nose region and their recognition performance tested under identification scenarios

As can be seen from Table 4.1, the three curves (L18L22, L20L24, L19L23) across the cheek/nose region shown in Figure 4.6 are the most significant curves of the 38 curves when selected by FSFS. The R_1RR of each selected curve is higher than the best curve selected on the nasal region, L9L1. Moreover, using all the three horizontal curves the R_1RR is 81.18%, which is comparable with the recognition performance of 21 selected nasal curves found in Section 4.5.3, despite the fact that 18 fewer curves are used.

4.5.4 Feature Selection Results of 113 Curves on the Cheek/Nose Region

FSFS has been successfully applied on the nasal (75) and cheek (38) curves selection, which shows the potential for using small subset of curves to produce higher recognition performance. In this section, the combination of the cheek/nose curves (75+38) is further evaluated by using the FSFS method. The number of points of each curve is resampled to 50 and the distance between L20 and L9 is set to 40 (pixels). The feature selection results from 113 curves on the cheek/nose region using the FSFS algorithm is shown in Figure 4.7. The selected curves and their corresponding R_1RRs are provided in Table 4.2.

Table 4.2: Curves selected by FSFS

	Curves selected by FSFS	**R_1RR**
1	L18L22	58.31%
2	L18L22, L20L24	76.98%
3	L18L22, L20L24, L9L1	82.65%
4	L18L22, L20L24, L9L1, L4L12	84.66%
5	L18L22, L20L24, L9L1, L4L12, L15L5	85.68%
6	L18L22, L20L24, L9L1, L4L12, L15L5, L19L23	86.24%
7	…L1L20	87.07%
8	…L1L20, L3L6	87.53%
9	…L1L20, L3L6, L16L10	87.72%
10	…L1L20, L3L6, L16L10, L14L6	87.98%
11	…L1L20, L3L6, L16L10, L14L6, L17L21	88.10%
12	…L1L20, L3L6, L16L10, L14L6, L17L21, L9L16	88.10%
13	…L1L20, L3L6, L16L10, L14L6, L17L21, L9L16, L1L14	88.36%
	L18L22, L19L23, L20L24, L9L1	**84.35%**

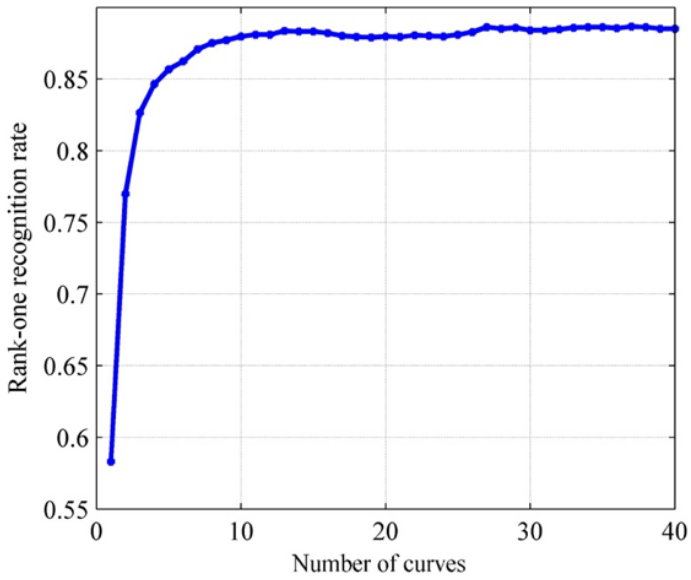


Figure 4.7: Rank one recognition rate against the number of curves selected by the FSFS algorithm using 113 curves on the cheek/nose region

To achieve the highest recognition performance, 13 curves were finally selected using FSFS to construct feature space. They are L18L22, L20L24, L9L1, L4L12, L15L5, L19L23, L16L18, L3L6, L16L10, L14L6, L17L21, L9L16 and L1L14. The R_1RR of those features is 88.36%, which shows that this feature combination can achieve good recognition performance. Therefore, using features from both the nasal region and its adjoining cheek region is beneficial for expression robustness. Finally, 4 curves are selected to form a low dimensionality rejector whose aim is to effectively and quickly eliminate a large number of ineligible candidate faces from the gallery.

4.6 Features Analysis

4.6.1 A Comparison between Proportional and Constant Distances for Cheek Region Boundary

For feature selection, three different proportional distances are tested: 2, 2.5 and 2.8 times the length of L5L9, respectively. For the constant distance, the horizontal length between the nose tip and cheek landmarks is set 40 mm and, for both methods, each curve is resampled to 50 points. The FSFS results for the 113 curves on the cheek/nose region presented in Figure 4.8 show that the recognition performance of the constant distance is always higher than the proportional distance, for all feature set sizes and proportional lengths. The highest R_1RR of 88.36% is obtained when 13 curves (L18L22, L20L24, L9L1, L4L12, L15L5, L19L23, L16L18, L3L6, L16L10, L14L6, L17L21, L9L16, and L1L14) are selected by constant distance. This

encouraging performance is significantly higher than that of [7] and shows the high potential of the cheek/nose region as a robust biometric.

The results in Figure 4.8 are interesting as, in theory, the proportional distances can preserve the whole face shape information and so can better represent the cheek landmarks for each person. However, in the experiments a constant distance performs much better than the proportional ones. The main reason for this is believed to be the instability of L5 and L13 landmarks, due to its sensitivity to noise. In comparison, the constant distance that only uses the location of nose tip is more reliable as the detection of nose tip is more accurate than those of the alar grooves and the root.

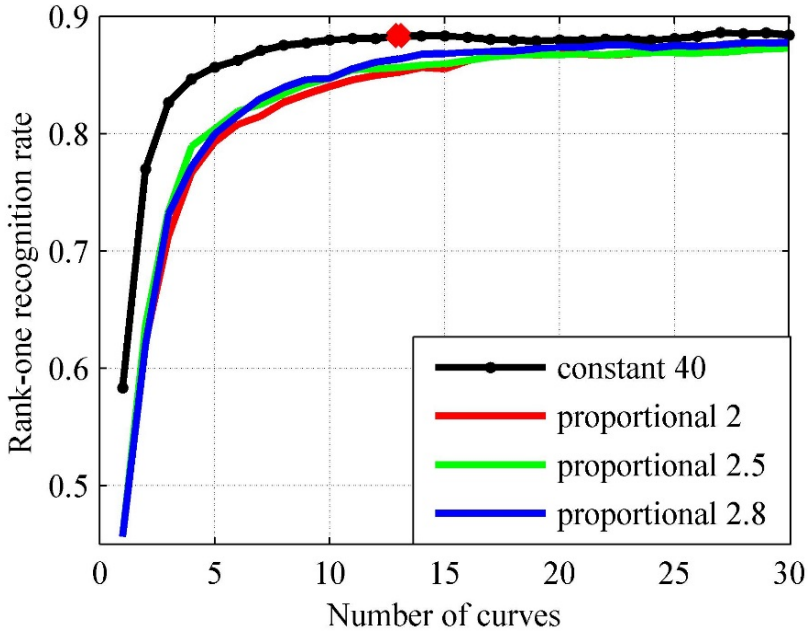


Figure 4.8: Rank one recognition rate against the number of curves on the cheek/nose region selected by the FSFS algorithm using a constant and proportional distances to locate landmarks L20 and L24.

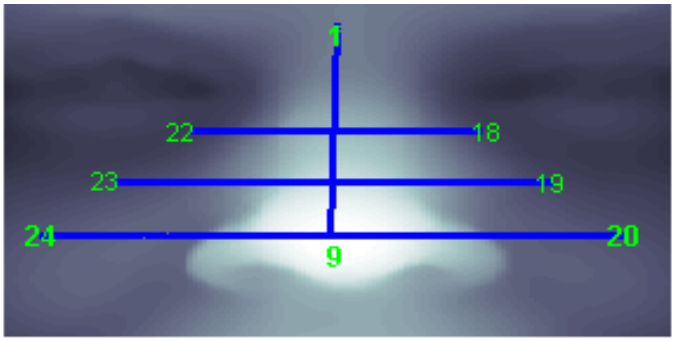


Figure 4.9: 4 curves extracted on the cheek/nose region.

The ideal scenario for pattern rejection is a high R_1RR achieved using a small number of features. In Figure 4.8, the highest R_1RR of 88.36% requires 13 curves, each of 50 points. If only the 4 curves shown in Figure 4.9 (L18L22, L19L23, L20L24, and L9L1) are selected the R_1RR drops to 84.35% but this performance is achieved using only 4×50 points. Therefore, using the cheek/nose region to form the feature set meets many of the requirements of a pattern rejector [136]. In the next three sections the use of different constant distances and the number of points on each curve are evaluated.

4.6.2 Requirements of Fixed Number of Points

According to a priori knowledge of the structure of the human face, these four curves in Figure 4.9 are generally represented with different lengths. Therefore, it is better to set different numbers of points on each curve to provide a more reasonable representation. To enable a comparison with fixed 50 points in the above experiments, the number of points are set to 50 (L9L1), 75 (L18L22), 100 (L19L23) and 125 (L20L24) respectively as the length of cheek curves is generally longer than curve L9L1. Figure 4.10 illustrates the CMC curves for both settings, 50-75-100-125 shown in blue and 50-50-50-50 that uses the constant number of points in each curve shown in red.

As can be seen from Figure 4.10, 50-50-50-50 is slightly better than 50-75-100-125 under identification scenario. Instead of using 50-75-100-125, different numbers have also been investigated in this experiment and the recognition performance is always lower than a constant number.

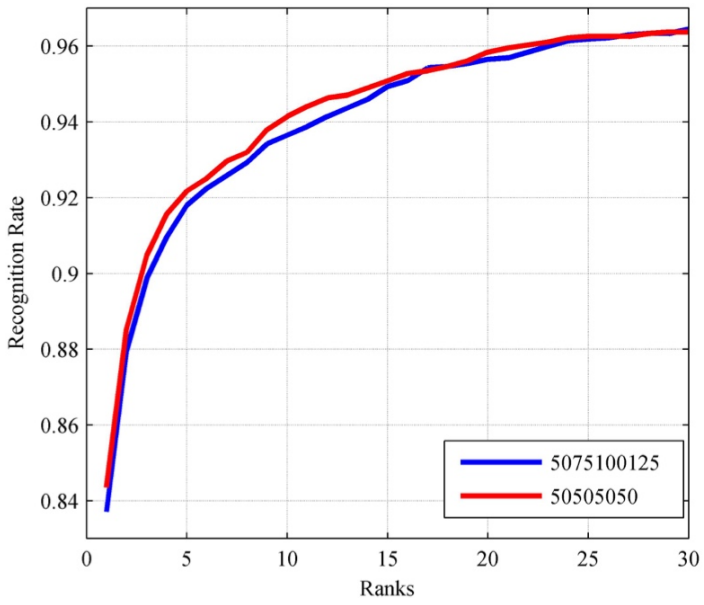


Figure 4.10: CMC curves for 50-75-100-125 and 50-50-50-50

4.6.3 The Choice of Constant Distances

The final lengths of the curves L18L22, L19L23 and L20L24 depend on the location of two cheek landmarks, which also directly determine how much of cheek region is used in the features. In the previous experiments, the constant distance was set to 40, as this enabled a fair comparison with the proportional distances. This is not necessarily the value that produces the best recognition results and here other values for the constant distance are investigated. As the cheek region is relatively constant over different expressions, using a large distance should be beneficial. However, larger distances have a higher probability of suffering from hair and hand occlusions. Therefore, to find the best performing distance, seven constant distances (30, 32.5, 35, 37.5, 40, 42.5 and 45) are defined and their recognition performance under the identification scenario is evaluated.

Figure 4.11 presents the Cumulative Match Characteristic (CMC) curves for the 7 different constant distances using the 'leave-one-out' identification scenario. Although the R_1RRs of all seven distances are similar, three distances (35, 37.5 and 40) outperform the shorter and longer ones. 37.5 and 40 produce higher recognition rates at lower ranks while 35 is better at higher ranks. This means that any distance between 35 and 40 will perform well and 40 is used in the following experiments.

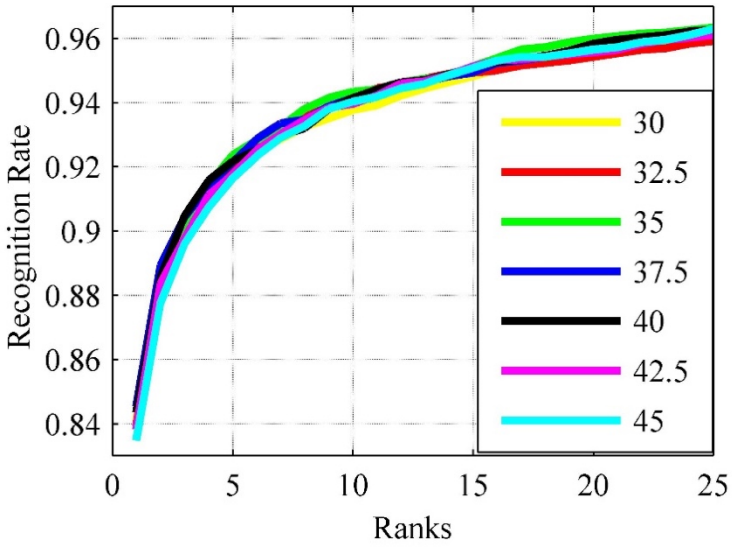


Figure 4.11 CMC curves for the 4 curves from Figure 4.9 with 50 points per curve using different constant distances to locate L20 and L24.

4.6.4 The Number of Points on Each Curve

During the feature selection phase, each curve was resampled to a fixed number of points. 50 points was chosen for the earlier experiments as initial results showed that it produced a very

similar R_1RR performance to 60 or 70 points but with fewer features to match. Generally, if the number is increased, more points will be extracted to construct the feature space which means more features are considered to represent the human face. However, the increase feature points do not necessary possess more discriminative information and such a high dimensionality also makes it hard to obtain a good classification performance [95, 96]. This fact is called "the curse of dimensionality", which means a large number of training samples should be added to classifier when the dimensionality increases. If the number decreases, less information will be used to evaluate the recognition performance. Therefore, 70 points per curve are chosen as the upper limit and 40 as the smallest one in the first tentative experiment. The results demonstrate that their recognition performances are approximately the same and simply increasing the size of the feature space cannot obtain high R_1RR.

In addition, the impact on the recognition performance of using fewer points is further investigated. The range used is from 5 to 45 points, in steps of 5 points, giving the 9 different CMC curves shown in Figure 4.12. The results in Figure 4.12 show that the recognition rate is maintained as the number of points reduces until a limit of 15 points is reached. After this, there is a small drop in performance for 10 points, particularly for ranks 12 and below, and a more significant decrease for 5 points. Therefore, using only 15 points can still produce a high recognition rate and, due to its low dimensionality, has much potential for pattern rejection.

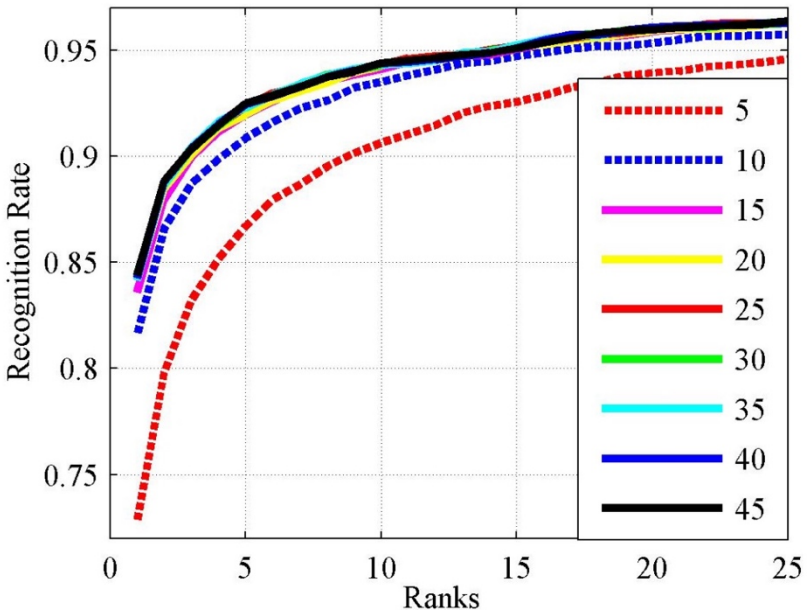


Figure 4.12: CMC curves for the 4 curves from Figure 4.9 using 5 to 45 points per curve. L20 and L24 are located using a constant distance of 40.

To further verify the best number of points for each curve, the classification performance of 4 curves (L18L22, L19L23, L20L24 and L9L1) is tested by KFA [137] with the polynomial kernel. The Bosphorus database is used for this experiment as it contains a wide repertoire of facial expressions. 1346 captures from 105 subjects are used as the training set and an additional 1300 captures from the same subjects provide the testing set. Ten different numbers from 5 and 50 are tested in this experiment and the resulting ROC and CMC curves are shown in Figure 4.13 and Figure 4.14, respectively. The R_1RR and EER for each number of points are given in Table 4.3. The EER is the most important parameter in the ROC curve and describes the trade-off between FAR and FRR, with a low EER indicating a good verification performance.

Table 4.3: R_1RR and EER for 5 to 50 points per curve using the KFA on the Bosphorus database.

No. Points	R_1RR	EER	No. Points	R_1RR	EER
5	70.96%	7.71%	30	92.22%	4.38%
10	90.29%	4.64%	35	91.83%	4.40%
15	**93.61%**	**3.63%**	40	91.29%	4.25%
20	93.76%	4.08%	45	90.83%	4.62%
25	92.30%	4.16%	50	91.06%	4.86%

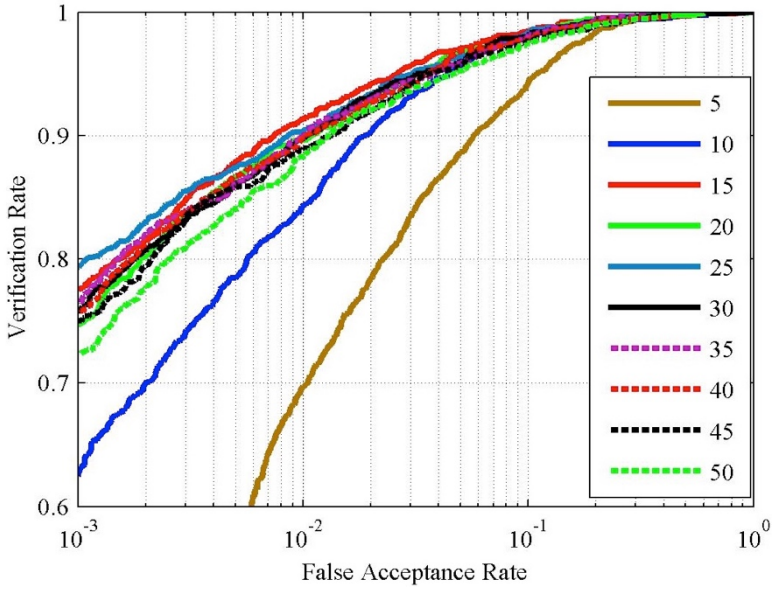


Figure 4.13: ROC curves for 5 to 50 points per curve using the KFA classifier on the Bosphorus database.

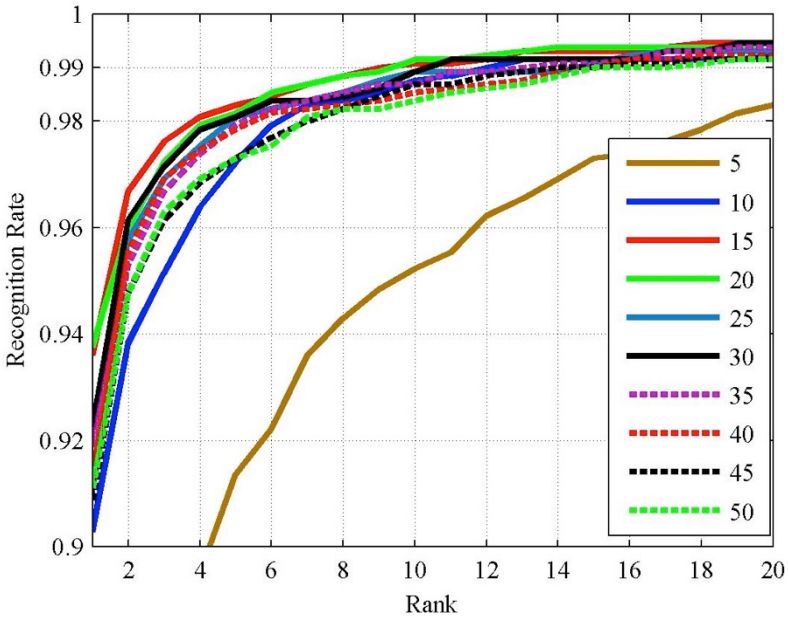


Figure 4.14: CMC curves for 5 to 50 points per curve using the KFA classifier on the Bosphorus database.

For the 'Leave-one-out' identification scenario, 15 is the smallest number which can still preserve a high R_1RR. As can be seen from Figure 4.13 and Figure 4.14, and Table 4.3, 15 points generally produces the best recognition performance for both scenarios, even though it only uses a total of 60 points across the four curves. For identification, using 15 and 20 points produces the highest recognition rate for all ranks. For verification, 15 points results in the lowest EER and clearly outperforms all higher or lower numbers. The R_1RR and EER results show that only 15 points per curve produces the best overall performance using only 60 points on the cheek/nose region. The classification performance of this rejector is evaluated in the next section.

4.7 Classification Performance Evaluation on the FRGC Database

To evaluate the classification performance of the rejector the FRGC database is used, as it has been reported to be the most challenging 3D face databases over varying expressions [111]. The Spring 2003 folder is used as the training set while the Fall 2003 and Spring 2004 folders provide the testing set. In addition, the sensitivity of the performance to the choice of constant distance to locate L20 and L24 is evaluated over a range distances. The ROC and CMC curves for 6 distances, ranging from 35 to 40, are shown in Figure 4.15 and Figure 4.16, respectively. Details of the parameters used and the resulting R_1RR and EER for the identification and verification scenarios, respectively, are provided in Table 4.4.

Table 4.4: Recognition performance using 4 curves with different constant distances

Scenarios	No. Points/Distance	R_1RR	EER
All vs All	15/35	86.08%	4.71%
	15/36	86.08%	4.36%
	15/37	86.23%	4.46%
	15/38	86.33%	4.17%
	15/39	86.28%	4.30%
	15/40	**86.33%**	**3.95%**

For both the identification and verification scenarios, the recognition performance using different constant distances show few variations. This shows that, within certain limits, the recognition performance is relatively insensitive to the constant distance used to select the landmarks. In general, the R_1RR increases and the EER decreases if larger distances are selected. Therefore, a constant distance of 40 with 15 points per curve is a good parameter choice for building the low-cost expression robust rejector for 3D face recognition.

The identification results in Figure 4.16 show that the recognition rate exceeds 98% at rank 20 which means the rejector can robustly reject a significant proportion of the gallery images using simple, low dimensional features from the cheek/nose region. The proposed rejector therefore has much potential to be applied in an efficient and robust classifier when combined with other features in a cascade classifier.

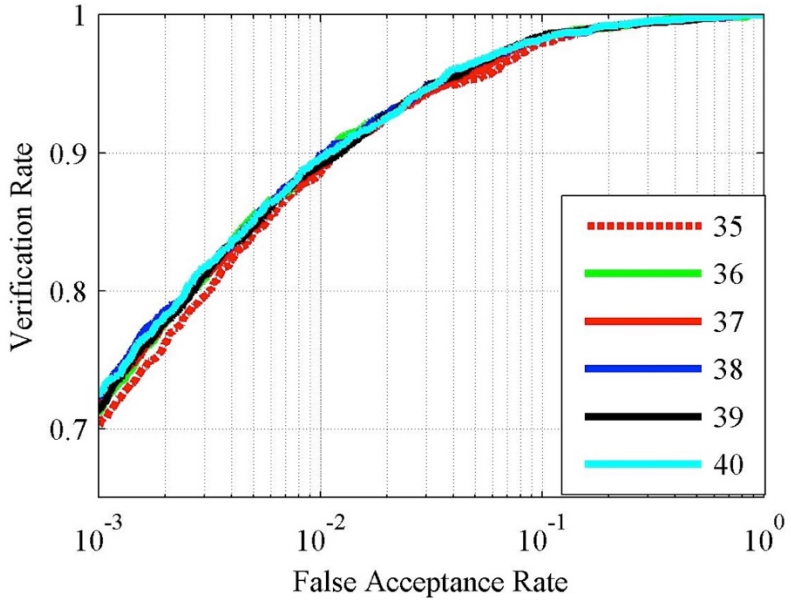


Figure 4.15: ROC curves produced by the KFA classifier on the FRGC database using different constant distances to locate L20 and L24.

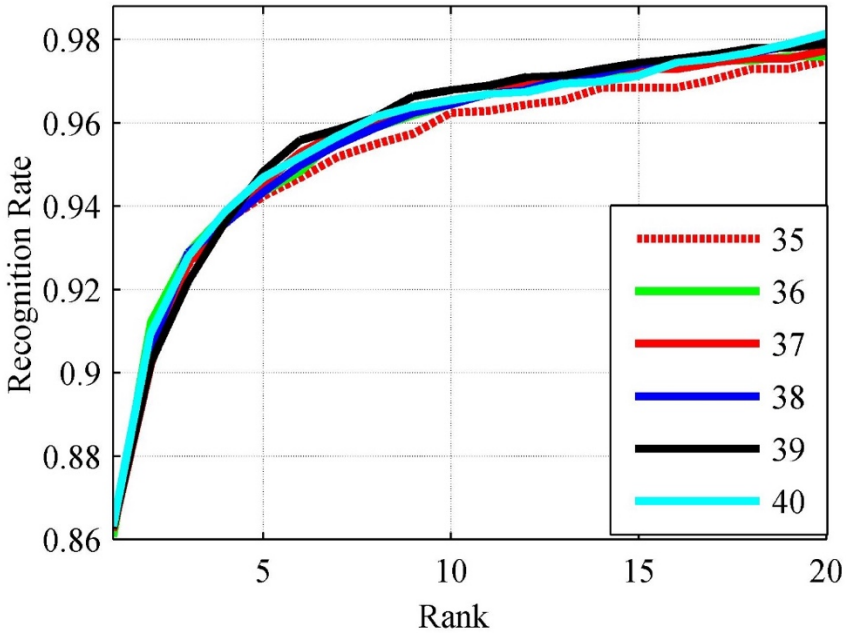


Figure 4.16: CMC curves produced by the KFA classifier on the FRGC database using different constant distances to locate L20 and L24.

To further evaluate the classification performance under different scenarios, the test sets Spring 2004 and Fall 2003, are further manually divided into neutral and non-neutral parts. During this process, it was observed that nearly all the captures in Spring 2003 (training) can be classified as neutral because only a few contain expressions and even these are low level expressions. Table 4.5 compares the R_1RR and the EER of the All vs All, All vs Neutral and All vs Non-neutral scenarios. In comparison, the EER value is much lower than the NCM algorithm of [7], which was ~8% for neutral and ~18% for non-neutral, respectively. The verification performance of proposed rejector is also much better than the reported EER of ~12% (neutral) and ~23% (non-neutral) for the multiple nose region algorithm of Chang *et al*. [4].

The proposed rejector shows a high classification performance, in particular for the non-neutral classes. Furthermore, the complexity is much lower than the NCM [7], as it only uses 4 curves with total of 60 points, compared to 28 curves with 1400 points. In addition, the location of these four curves mainly depends on the nose tip, resulting in more stable curves.

For an additional comparison, the proposed rejector is compared to that of Mian *et al.* [5]. Although [5] combined both 2D (SIFT) and 3D (SFR) features to form the rejector, to provide a comparable classification performance only the 3D SFR results are used here. A comparison of the effectiveness of proposed rejector and SFR-based rejector of [5] is shown in Figure 4.17 and Table 4.6. At the 98% verification rate the $Eff(\psi)$ of proposed rejector for the whole gallery

is 8.4%, such that all bar 39 of the 466 faces are remained for further classification. This is much lower than SFR-based rejector's result of 38% and shows that the proposed rejector can quickly eliminate 92% of the gallery faces. When the gallery is divided into neutral and non-neutral, the $Eff(\psi)$ is still significantly lower for the proposed technique than for the SFR-based rejector for both scenarios.

Table 4.5: The R_1RRs and EERs of proposed pattern rejector under different scenarios, including "All vs All", "All vs Neutral", "All vs Non-neutral".

No. Points/Distance	Scenarios	R_1RR	EER
15/40	All vs All	86.33%	3.95%
	All vs Neutral	92.68%	2.73%
	All vs Non-neutral	75.32%	5.78%

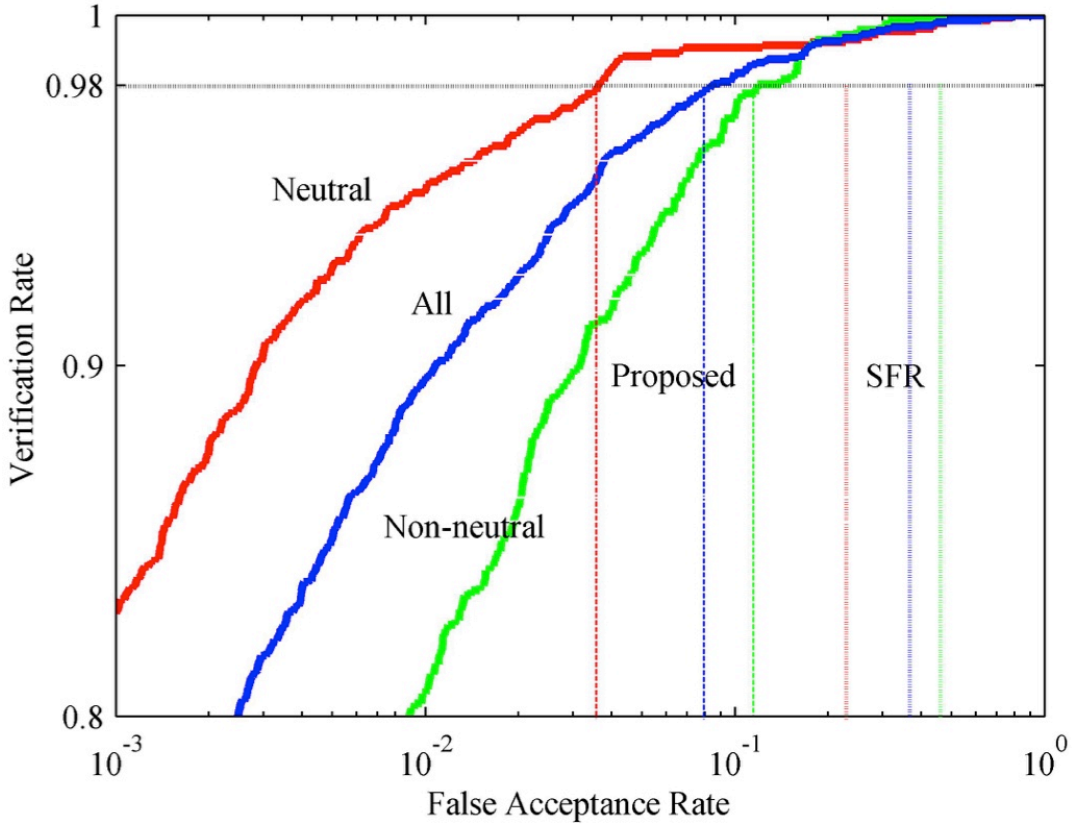


Figure 4.17: Rejection classification results at the 98% verification rate compared to SFR in [5].

Table 4.6: Effectiveness evaluation of the proposed rejectors

Methods	*Eff*(ψ)		
	All	**Neutral**	**Non-neutral**
4 curves	0.084	0.036	0.118
SFR (Mian *et al.* [5])	~0.38	~0.24	~0.47

4.8 Radial and Root Curves Evaluations

Instead of using the horizontal curves, some other kind of curves, radial and root, are also investigated on the detected nasal and adjoining cheek regions. After the preprocessing and normalization of the 3D captures, 16 landmarks are detected on the perimeter of a triangular structure containing the nose region, as shown in Figure 4.18(a). The first set of curves are the nine radial curves shown in Figure 4.18(c), connecting the nose tip to the landmarks on the edge of the triangular region. Compared to the radial curves in [49, 80], the nose tip is automatically detected in the preprocessing stage and the curves only contain the nose and adjoining cheek regions, which are relatively stable on the human face and are not affected by natural occlusions.

Similarly, nine root curves from the tip to the lower boundary of the triangular structure and evenly distributed on the nose and adjoining cheek region can be defined, as shown in Figure 4.18(d). Previous work has found that resampling curves on the nose surface to 15 points per curve produced the best recognition performance [138]. As the radial and root curves are shorter than the horizontal curves used in [138], and smaller feature sets are generally desirable, the R_1RR and EER performance using 10 and 15 points per curve for both the radial and root curves are evaluated and the recognition results presented in Table 4.7 and Table 4.8, respectively. For the radial curves, Table 4.7 shows that using 10 or 15 points produces similar R_1RR and EER for the three scenarios evaluated for the FRGC database. However, when tested on the Bosphorus database using 10 points performed better than using 15. For the root curves, the recognition performance in the Table 4.8 demonstrates that 15 points per curve is marginally better than 10 which may be because the root curves are longer than the radial curves.

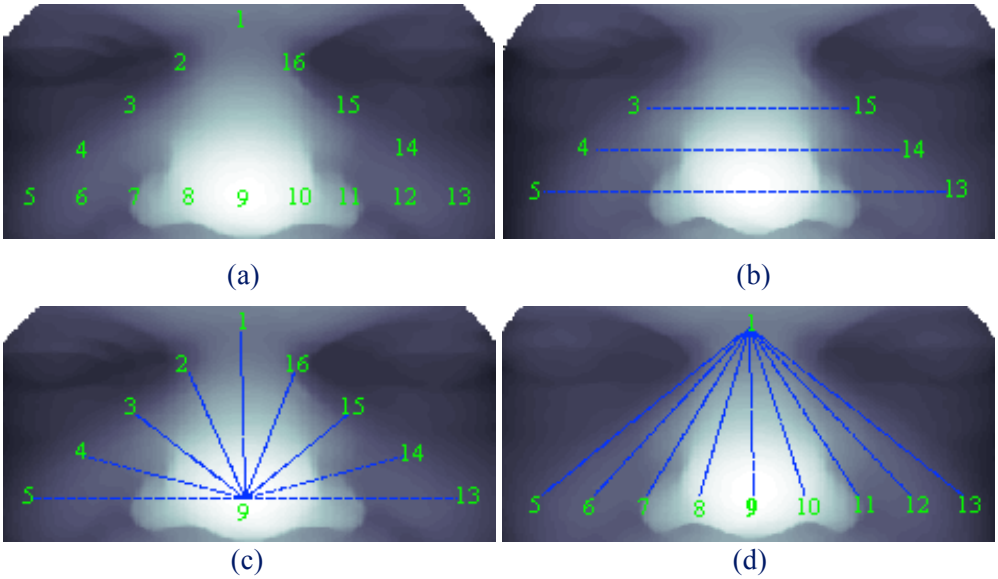


Figure 4.18: Landmarks and curves extracted on the cheek/nose region. (a) 16 landmarks; (b) three horizontal curves in [138]; (c) nine radial curves from the nose tip; (d) nine curves from the nose root.

Table 4.7: Recognition performance of the radial curves tested under different scenarios using the Bosphorus and FRGC databases.

	Scenarios	**10 points**		**15 points**	
		R_1RR	EER	R_1RR	EER
Bosphorus[13]	All vs All	94.07%	3.49%	92.30%	3.70%
FRGC [12]	All vs All	88.88%	4.46%	89.18%	4.40%
	All vs Neutral	93.69%	2.71%	94.31%	2.72%
	All vs Non-neutral	80.22%	5.75%	79.94%	6.17%
Features set size for each capture		10×9 = 90		15×9 = 135	

Table 4.8: Recognition performance of the root curves tested under different scenarios using the Bosphorus and FRGC databases.

	Scenarios	**10 points**		**15 points**	
		R_1RR	EER	R_1RR	EER
Bosphorus[13]	All vs All	91.60%	3.86%	91.60%	3.69%
FRGC [12]	All vs All	86.68%	4.66%	87.23%	4.75%
	All vs Neutral	91.82%	3.12%	92.13%	3.11%
	All vs Non-neutral	77.42%	6.14%	78.40%	6.31%
Features set size for each capture		10×9 = 90		15×9 = 135	

4.9 Conclusion

To improve the classification performance under different expressions and over a large number of classes, a new 3D face recognition algorithm is proposed in this chapter by matching the curves extracted on the nasal and adjoining cheek regions. After preprocessing, consisting of face cropping, denoising and alignment, 24 landmarks and 113 curves on the cheek/nose region are generated to form the feature set. Using the Bosphorus 3D database, FSFS is applied to select the most effective 4 curves over various expressions.

The key landmarks are those located on the cheeks, either side of the nose, and using a constant horizontal distance from the nose tip was found to perform better during this stage than extrapolating a proportional distance from the nose tip to the alar; this was due to the inconsistencies in the alar landmarks. Performance evaluation using the FRGC database further showed that, for a constant distance on 40 mm, the best recognition rates are achieved using 15 points per curve, giving only 60 points in total. This low dimensionality feature set can be used as a pattern rejection classifier for 3D face recognition.

In addition to the horizontal curves, some other kind of curves, radial and root, are also evaluated on the detected nasal and adjoining cheek regions, using 10 or 15 points per curve for both the radial and root curves. Compared to the root curves, the radial curves produce higher recognition performance in both scenarios, especially when the expressions occur.

Chapter 5 Matching Multi-Component Local Shape Descriptors

5.1 Introduction

Overcoming variations in human expressions, poses and occlusions are some of the most challenging topics in face recognition. Using 3D data facilitates the correction of pose and can help address the problems caused by occlusions. However, variations in expression which lead to muscle movements and deform the face surface still present challenges to recognition systems using 3D data. One way to address these problems is to select relatively stable structures and patches on the facial surface and use these to explore its discriminative features to design expression invariant face recognition algorithms. The use of the 'cheek/nose' region in Chapter 4, defined by a triangular structure containing the nasal and adjoining cheeks, showed the potential of extending the nasal region to include adjoining parts of the cheeks for expression robust 3D face recognition. The use of the nasal and adjoining cheek region supports the hypothesis that they are relatively stable under expressions and less affected by occlusions.

To explore the recognition potential of local patches, many works employed some local descriptor to extract features such as Local Binary Patterns [115]. Li *et al*. successfully used multi-scale and multi-component local normal patterns to find expression robust patches on the whole 3D face, showing the discriminatory power of both Local Binary Patterns and surface normals [6]. However, their use of three scales and three components on the whole face results in high dimensional feature set for matching and can also be adversely affected by occlusions and expression variations. Therefore, exploring the local normal features on relatively rigid regions of the face is an attractive area of investigation.

Most 3D face recognition algorithms are focused on using depth information to select features and produce good recognition performance. In addition to depth, the surface normals of each point determine the orientation of the surface and contain the information on local shape variations. For example, Zafeiriou *et al*. obtained the surface normals information from photometric stereo captures and used it as a normal face for face recognition [9]. In comparison, Li *et al*. used the surface normals calculated from 3D point cloud data which is captured from laser scanner and demonstrated a good recognition performance [6]. These works provide the

motivation for the fusion of depth and surface normals data to exploit more discriminative features for 3D face recognition.

This chapter further investigates the use of the nasal and immediately adjoining cheek regions as a biometric. It proposes a novel local shape descriptor to find expression robust discriminative features on both depth and surface normals maps. This descriptor is calculated from the points selected from a set of three horizontal lines and their eight neighbours at a fixed separation. Evaluating the features over various scales demonstrates that this local shape descriptor provides a novel way to represent the shape changes on the face surface and produce a better recognition performance than using depth or surface normal values independently. In addition, a further analysis of the discriminatory power using 24 patches on the nose and adjoining cheek regions is also evaluated. A discriminatory power map is produced, showing the potential of the region immediately adjoining the nose as a biometric. On the basis of the resulting discriminative map of the depth and three normal components, an extension of this new local shape descriptor is applied to the nine patches on the selected region. The recognition performance demonstrates that features extracted on this combined region by the proposed local shape descriptor can be developed as an expression robust 3D face recognition algorithm.

5.2 3D Landmarks Localization

After the preprocessing and normalization of the 3D captures, 16 landmarks are detected on the perimeter of a triangular area containing the nasal region, as shown in Figure 5.1. Surface features are extracted to first defined planes connecting any two landmarks, which are perpendicular to the *x*-*y* plane. This results in a series of curves on the face surface [7, 138]. Feature sets can then be found by normalizing the curves and then resampling them to a fixed number of points using bicubic interpolation. Gao *et al*. found that using 15 points from horizontal lines was an effective approach to capturing the discriminative features [138].

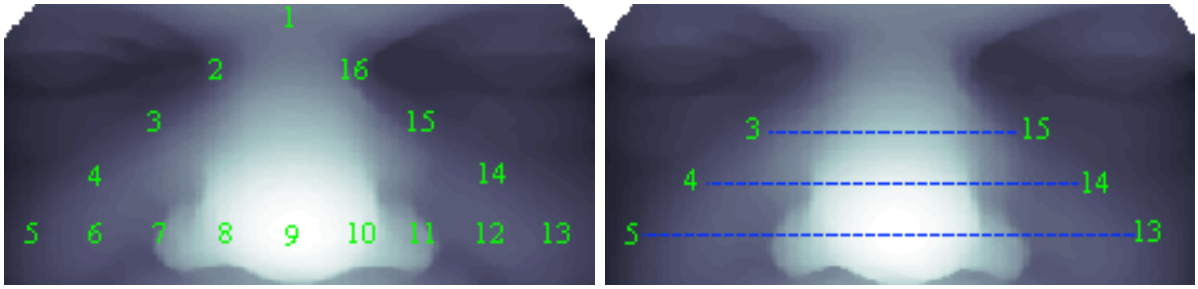


Figure 5.1: 16 Landmarks and 3 horizontal curves extracted on the cheek/nose region.

5.3 An Expression Robust Multi-Component Local Shape Descriptor

5.3.1 Building Local Shape Descriptors by Horizontal Curves

As an alternative to using the direct depth values of points from curves on and around the nasal region, this chapter presents a novel representation for describing the shape changes by computing depth differences, in a similar manner to Local Binary Patterns [9]. From the points extracted on the three horizontal curves shown in Figure 5.1, a series of windows (the white quadrilaterals in Figure 5.2) are used to compute the depth differences between the central points (red points) on the middle curve and the surrounding points (white points) on the top, middle and bottom curves. Specifically, all the even points on the middle curve are considered as the central points of 3×3 windows and the depth differences with their eight neighbours are stored as $D_w = [dd_1, dd_2, dd_3, dd_4, dd_5, dd_6, dd_7, dd_8]$. Therefore, the feature vector for each capture can be defined as $(D_{w2}, D_{w4}, D_{w6}, \ldots, D_{w(i-1)})$, in which i is the number of points on each curve. In the experiments, i is set to 15 so there are 15 points on each curve and a total of 7 windows for calculation, giving a feature set size of 8×7 = 56. For comparison, a feature set of size 15×3 = 45, consisting of the depth values at the 15 points on each curve is used.

The recognition performance is measured by the EER for verification and the R_1RR for identification. KFA [139] with the polynomial kernel is used to calculate the similarity between the gallery and probe captures and the cosine distance is used in the feature matching. A comparison of recognition results between this new local shape descriptor and direct depth features is shown in Table 5.1. The results show that using the depth differences performs better for all scenarios. In particular, they produce a R_1RR that is about 9% higher than the linear depth features for the "All vs Non-neutral" scenario. This demonstrates that this local shape descriptor has the potential to produce a better recognition performance than using the depth values directly, especially when the captures contain different kinds of expressions.

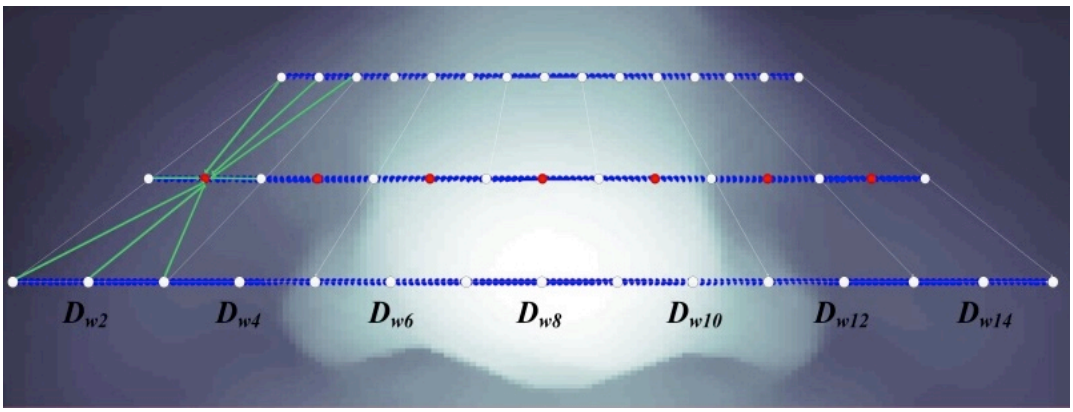


Figure 5.2: Depth difference calculation using three horizontal curves.

Table 5.1: A comparison of the R_1RRs and EERs between the depth difference and direct depth features extracted from the horizontal curves shown in Figure 5.2. The captures from Bosphorus and experiment 3 in FRGC are used for recognition performances evaluation.

	Scenarios	**Depth Difference**		**Direct Depth**	
		R_1RR	**EER**	**R_1RR**	**EER**
Bosphorus [13]	All vs All	88.98%	4.93%	85.82%	5.10%
Experiment 3 in FRGC [12]	All vs All	83.53%	5.34%	79.32%	5.32%
	All vs Neutral	90.50%	2.88%	89.02%	3.42%
	All vs Non-neutral	70.91%	7.02%	61.85%	8.56%
Features size of each capture		8×7 = 56		15×3 = 45	

5.3.2 Local Depth and Normals Features Extraction on the Nose and Adjoining Cheeks

The results in the previous section support the conclusions of previous researchers, that the nose and adjoining cheek regions provide a good source of discriminative features [8, 80, 138]. In addition, in comparison to using the depth directly, using the proposed local shape descriptor to build the feature vector by calculating the depth differences is an effective and efficient method to describe the shape of the nasal region. In the previous section, the depth differences were calculated over a relatively large scale and in this section the discriminative features for expression invariant 3D face recognition are further explored by computing the depth difference more locally on the nose and adjoining parts of the cheek.

With reference to Figure 5.3, the local depth descriptor is found using the points on the three horizontal curves shown in Figure 5.2. Again, the curves are resampled to 15 points using bicubic interpolation and all the even points on the curves (marked as red points in Figure 5.3) are taken as the central points of the 3×3 windows for which the depth differences are found. For each central point, the eight neighbouring points are given by the left and right points on the same curve and the three points directly above and below from curves at a vertical separation D, shown in blue in Figure 5.3. D is measured in pixels at the resolution of the original 2D maps produced from the 3D point cloud. To allow the scale for different regions to be explored independently, the vertical distances D_1, D_2 and D_3 are used for the top, middle and bottom curves, respectively. For the seven central points on each of the three curves the resulting depth features produce a feature vector of size 8×7×3 = 168.

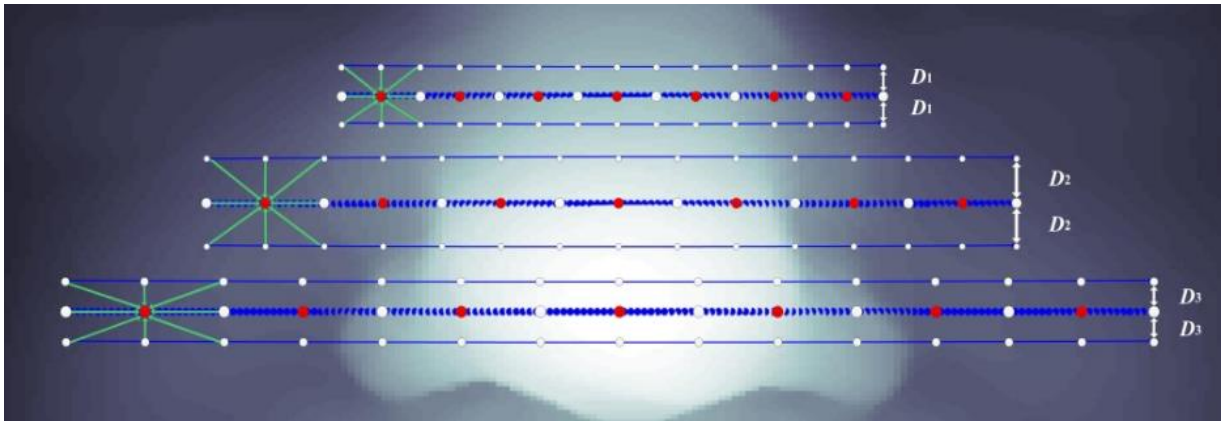
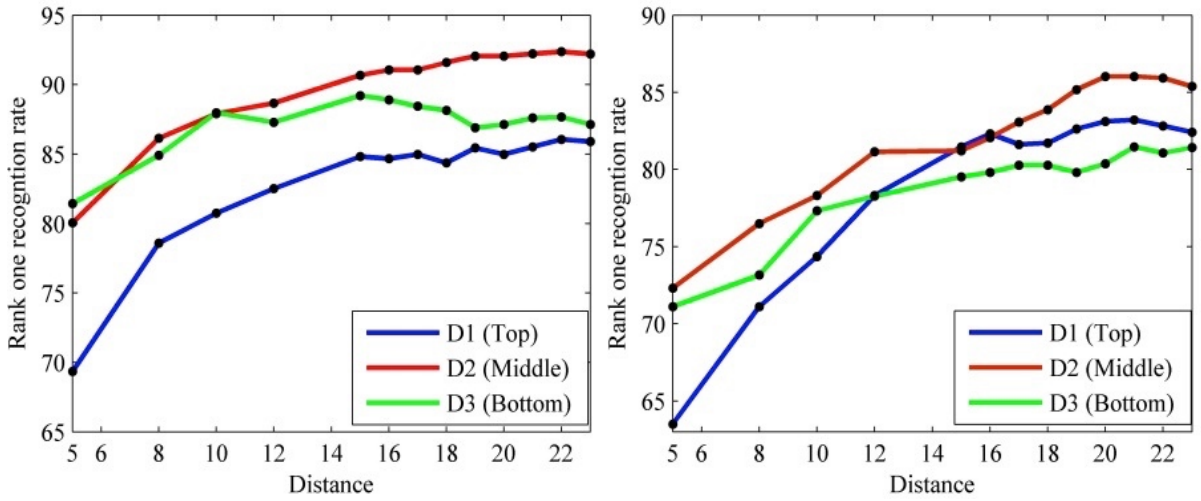


Figure 5.3: Local features extraction by calculating depth difference around three horizontal curves.



(a) FRGC (b) Bosphorus

Figure 5.4: Rank one recognition rate of depth difference extracting around the three horizontal curves under different distances on the FRGC database.

Figure 5.4 plots the R_1RRs achieved by the depth descriptor for the top, middle and bottom curves using different values of distance D for both the FRGC and Bosphorus databases. In general, the depth descriptors extracted from each horizontal curve produce better results as the vertical separation D increases. Depth differences from the middle curve produce a higher R_1RR than those from the top and bottom curves, yielding the highest R_1RR of 86.03% when $D_2 = 20$ for the FRGC database and 92.37% when $D_2 = 22$ for the Bosphorus database. The R_1RR of depth differences from the top curve increases up to $D_1 \approx 15$ and becomes more stable afterwards. In contrast, the R_1RR of depth differences from the bottom curve increase more slowly and when $D_3 \geq 10$ it becomes relatively constant, with some fluctuations.

An improved performance can be achieved by combining the depth differences from all three curves. However, as the vertical separation D increases the depth differences from the top and bottom curves are more affected by natural occlusions (e.g. glasses and hands) and shape changes caused by expression, respectively, and so simply combining the best performing distances from each curve will not necessarily produce the best combined performance. An evaluation of the combined depth difference features on the FRGC and Bosphorus databases using different distances for D_1, D_2 and D_3 was undertaken and the best performance was found to occur when D_1 and D_3 are ~15 and D_2 is ~20.

Using these parameters, the recognition performance is evaluated under both the identification and verification scenarios and the results are presented in Table 5.2. The performance for captures with expression variations is significantly better than those achieved using the large scale depth difference (see Table 5.1), albeit with a larger feature set. In addition to depth difference calculation, the local difference descriptor is also applied to the surface normals maps to find the discriminative features from the SN_x, SN_y and SN_z components, producing another three different representations of this region. The scale of the windows for the normals was the same as that used for the depth and the recognition performances for both databases are illustrated in Table 5.2.

Surface normal features outperform the depth ones on the Bosphorus database, which contains many variations in expression, and also for the FRGC "All vs Non-neutral" scenario. Results also show that the R_1RR for the Bosphorus database is much higher than for FRGC experiment 3. This is most likely due to the noisy training captures in the Spring 2003 folder.

Table 5.2: Recognition performance of the depth and surface normals differences extracted around the three curves shown in Figure 5.3.

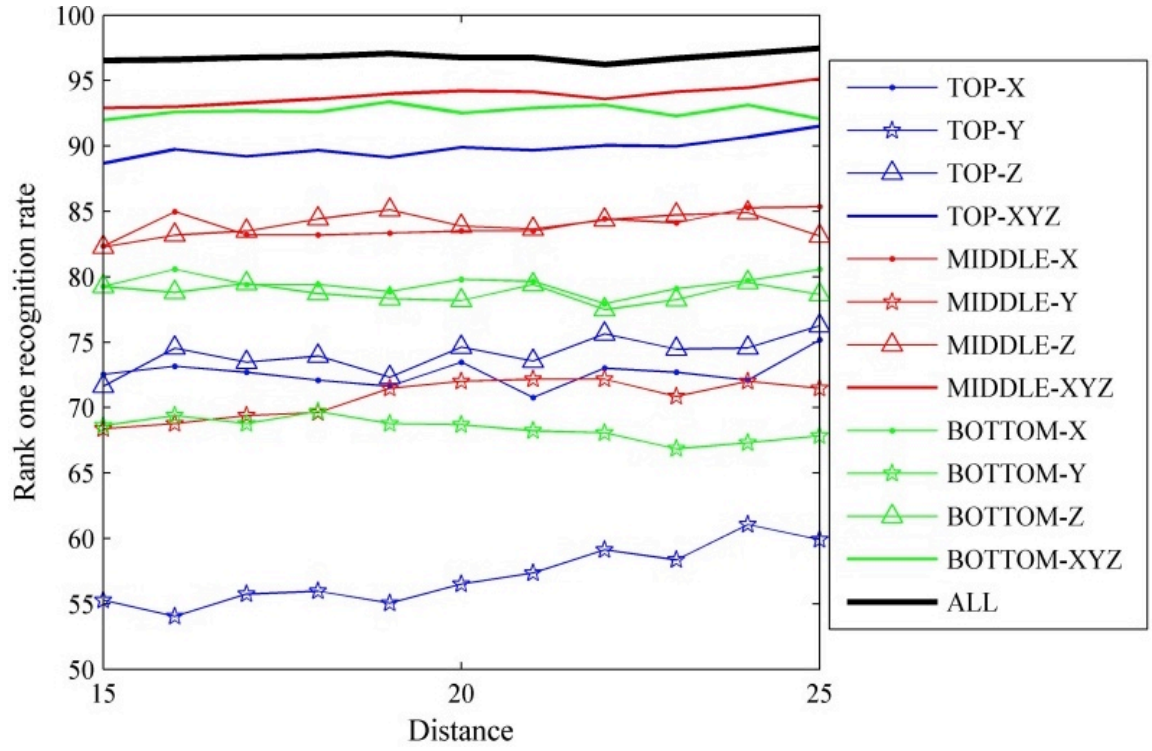
	Scenarios	**Depth Features**		**Normal Features**	
		R_1RR	**EER**	**R_1RR**	**EER**
Bosphorus [13]	All vs All	94.99%	3.47%	97.76%	1.32%
Experiment 3 in FRGC [12]	All vs All	89.63%	4.51%	89.08%	3.90%
	All vs Neutral	93.38%	3.03%	91.67%	3.11%
	All vs Non-neutral	82.89%	5.88%	84.13%	4.91%
Features size of each capture		8×7×3 = 168		8×7×3×3= 504	

Table 5.3: R_1RRs of features extracted from three components of surface normals (SN_x, SN_y and SN_z) and three horizontal regions (Top, Middle and Bottom). The Bosphorus, experiment 3 setting in FRGC and FRGC v.2 captures are used for recognition performances evaluation.

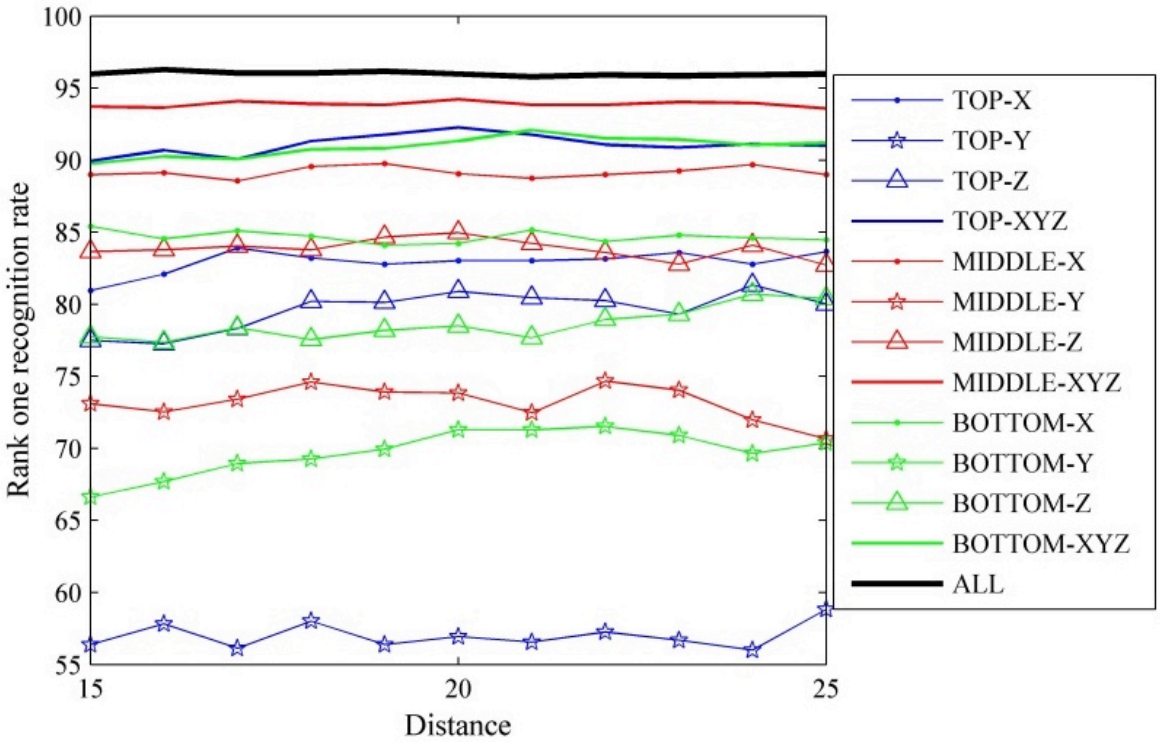
	Components	**Size**	**Bosphorus**[13]	**Experiment 3 in FRGC** [12]	**FRGC v.2** [12]
Surface Normal	SN_x	168	94.60%	83.62%	94.72%
	SN_y	168	90.28%	78.42%	90.19%
	SN_z	168	94.83%	83.78%	92.77%
	Top	168	91.52%	81.87%	92.26%
	Middle	168	95.14%	83.93%	94.21%
	Bottom	168	93.37%	79.36%	92.08%
	All	504	97.76%	89.08%	96.29%
Depth	Top	56	84.81%	83.22%	84.16%
	Middle	56	92.06%	86.03%	91.39%
	Bottom	56	87.12%	81.47%	87.12%
	All	168	94.99%	89.63%	96.17%

In addition to the analysis of the recognition performance for the depth component in Figure 5.3, some additional identification evaluations are used to further explore the discriminatory power of different locations on the selected region (Top, Middle and Bottom) and three surface normal components (SN_x, SN_y and SN_z) using a fixed scale, see Table 5.3. The R_1RR results are for the Bosphorus database, experiment 3 of FRGC and the FRGC v.2 databases. In general, the results in Table 5.3 show that the SN_x and SN_z possess more discriminative features than SN_y and produce a performance that is ~4% better. For the horizontal regions of cheek/nose region, both the depth and normal features extracted from Middle part outperform those from the Top and Bottom parts.

An additional evaluation on the relationship between the R_1RR and different scales is shown in Figure 5.5, where horizontal curves with separations ranging from 15 to 25 pixels are investigated. The results show that the R_1RR remains stable when different separations are selected, proving that the proposed local shape descriptor is scale robust to some extent. Also, the SN_x, SN_z and Middle components always produce better R_1RRs, regardless of which scale is selected.



(a) Bosphorus



(b) FRGC v.2

Figure 5.5: R_1RR of features extracting from horizontal curves with different separation distances using three components of surface normals (SN_x, SN_y and SN_z) and three horizontal parts (Top, Middle and Bottom) of the nose and adjoining regions on the Bosphorus and FRGC v.2 databases.

The recognition results for the whole FRGC database shown in the Table 5.2 and Table 5.3 reveal that surface normals are much more sensitive to the noise caused by lighting and the capture device. Training data from noisy folder, Spring 2003, affects the within-class similarity and between-class dissimilarity and so cannot reliably be used to select the most discriminative normal features. To investigate this, Figure 5.5 evaluates the surface normal features on the FRGC v.2 database, 466 subjects with 4007 captures in total, and presents the R_1RR of features selected from three components (SN_x, SN_y and SN_z) and three parts (Top, Middle and Bottom) using different distances. The recognition performance achieved is quite similar to that of the Bosphorus database, showing the adverse effect of the noisy FRGC Spring 2003 captures on the recognition performance.

5.3.3 Recognition Performance Comparison

For classification and matching, the KFA classifier with the cosine distance produces a R_1RR of 97.76% and an EER of 1.32% for the Bosphorus dataset, which is competitive with methods whose algorithms are more sophisticated and require a larger feature set. Comparison with other research that considers the nasal region and its environs, see Table 5.4 and Table 5.5, shows that the proposed method has much potential.

For verification, using the FRGC database (Experiment 3), the EER for both depth and surface normals was found to be ~3% for neutral and ~5% for non-neutral which compares favourably with the results of Chang *et al*. [4] (12% and 23%, respectively) and Emambakhsh *et al*. [7] (8% and 18%, respectively). For identification, compared to the algorithms using both the selected region and whole face in Table 5.5 and Table 5.6, the combined surface normals results exceed 96.29% and 97.76% on the FRGC v.2 and Bosphorus databases, respectively, with a relatively small and simple feature set which is within 2% of the state-of-the-art performance using the whole face.

Table 5.4: A comparison of EERs of features extracted from the nasal region and its environs on the FRGC database (Experiment 3).

Algorithm	Matching	Neutral	Non-neutral
Chang *et al.* [4]	ICP	0.12	0.23
Emambakhsh *et al.* [7]	KFA-Poly	0.08	0.18
Gao *et al.* [138]	KFA-Poly	0.03	0.06
Proposed method	KFA-Poly	0.03	0.05

Table 5.5: A comparison of R_1RRs of features extracted from the nasal region and its environs on the FRGC (Experiment 3) and Bosphorus databases.

Algorithm	FRGC		Bosphorus
	Neutral	Non-neutral	
Chang *et al.* [4]	96.6%	82.7%	
Emambakhsh *et al.* [7]	90.87%	81.61%	97.44%
Gao *et al.* [138]	92.68%	75.32%	93.61%
Proposed method	93.38%	84.13%	97.76%

Table 5.6: A comparison of R_1RRs of different recognition algorithms tested on the FRGC and Bosphorus databases.

	Algorithm	FRGC		Bosphorus
		Experiment 3	v.2	
Nasal and its environs	**Gao *et al.*** [138]	86.33%		93.61%
	Wang *et al.* [8]		~92%	
	This work	89.63%	96.29%	97.76%
The whole face	**Wang *et al.*** [8]		98.3%	
	Li *et al.* [6]		96.3%	95.4%
	Kakadiaris *et al.* [140]		97.9%	98.2%
	Ballihi *et al.* [80]		98.02%	
	Drira *et al.* [49]		97%	
	Faltemier *et al.* [141]		97.2%	

Using a simple descriptor to extract features results in a very small feature set, which has much potential for real time automatic face recognition system as all the preprocessing parts (face cropping, alignment and landmark localization) are fully automated and no feature selection is required after the feature extraction.

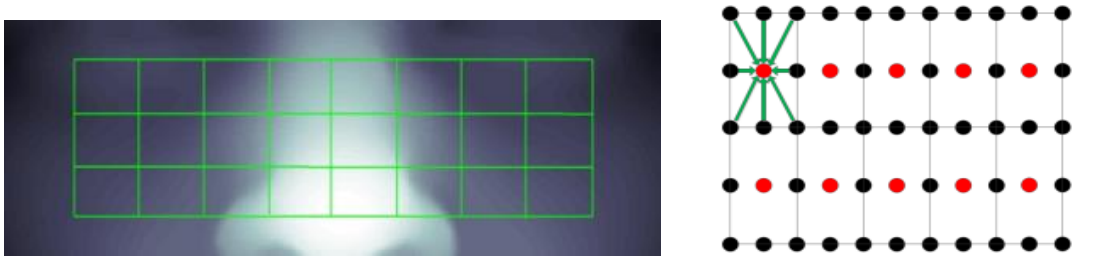
5.4 Local Patches Evaluation on the Nasal and Its Environs

The proposed local shape descriptor demonstrates good potential for extracting expression robust features on a larger scale with a small sized feature set, using the nose and adjoining cheek regions as an expression robust biometric. It is therefore interesting to use this descriptor to explore the discriminatory power more locally.

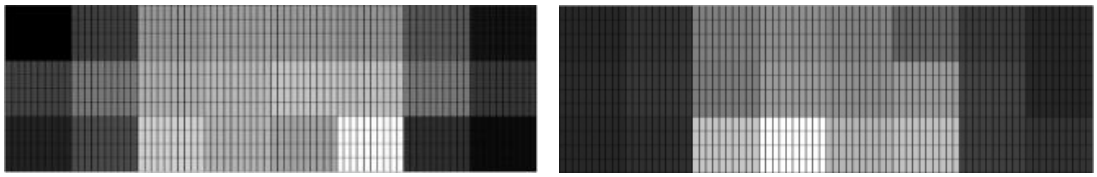
5.4.1 Discriminatory Power Evaluation on Local Patches

The nasal and adjoining cheek regions shown in Figure 5.6(a) are used to further investigate the discriminative features for expression robust face recognition, which is determined by two cheek landmarks, and the top and bottom horizontal curves shown in Figure 5.1. This is an extension of the region discussed in Section 5.3, which is further divided into 24 small local patches, each of which is resampled to 5×11 pixels, as illustrated in Figure 5.6(b). Calculating the local shape descriptor proposed in Section 5.3 builds the feature set from both the depth and surface normals components.

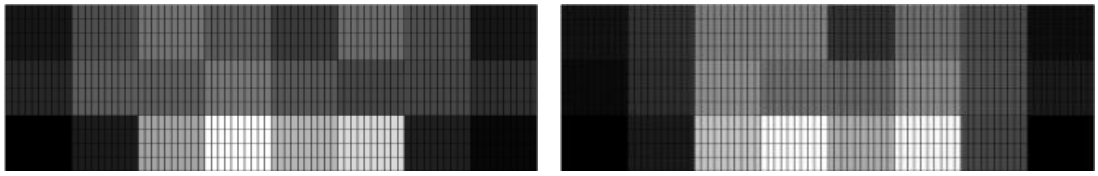
The R_1RR of each patch is obtained using the KFA classifier and cosine distance. To illustrate the discriminatory power, the R_1RRs for the patches are presented as grayscale images in Figure 5.6(c) - (f), where a brighter patch represents a higher R_1RR. Results show that the R_1RR of patches on the nasal region and its adjoining cheek regions is always higher than those on the cheek bone region, for all four components.



(a) 24 patches (b) Proposed Local shape descriptor



(c) Depth (d) SN_x



(e) SN_y (f) SN_z

Figure 5.6: The patch weights corresponding to depth image and three normal components images learned from the Bosphorus database. The weights are presented as grayscale values, where brighter values denote larger weights.

Furthermore, it is more interesting to observe from Figure 5.6 that those patches located on the boundary of the nasal and adjoining cheek regions perform better than those only wholly in the nasal region which shows the potential of the cheek regions immediately adjoining the nose. The features were also evaluated on the FRGC v.2 database, producing similar recognition performances and discriminatory power distribution for the depth and three normal components to that shown in Figure 5.6. This analysis shows that the grayscale images in Figure 5.6 can be considered to be a good representation of the discriminatory power on the nasal and adjoining cheek regions.

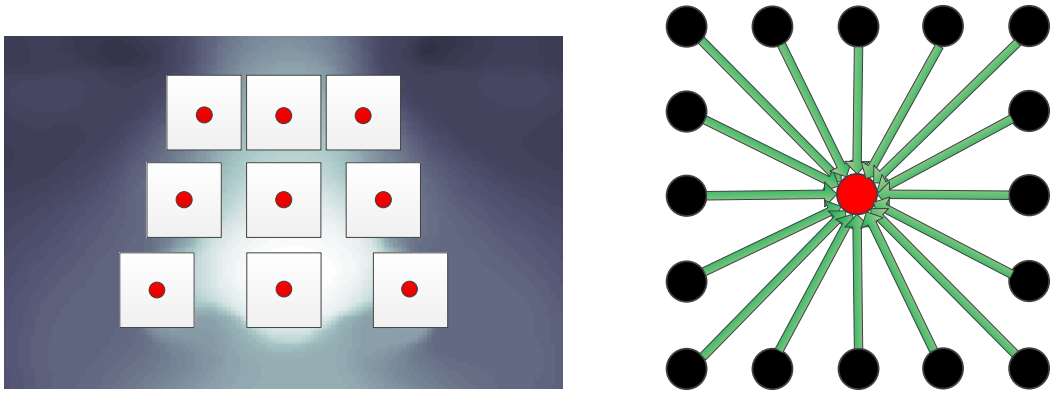
5.4.2 Square Local Shape Descriptor

As can be seen from Figure 5.6, the proposed local shape descriptor produces better recognition performance on the nasal and its neighbouring patches. Therefore, to further investigate the proposed descriptor and extract the discriminative features, nine landmarks (red points) shown in Figure 5.7(a) are localized accurately to define the nasal and its environs region. The two nasal alar landmarks are found by intersecting a set of planes which are perpendicular to the *x-y* plane. This results in a set of curves and the alar landmarks are located at the minimum of the curves' maxima [7]. The vertical middle three landmarks are determined by the location of saddle and tip and are evenly distributed on the line from tip to saddle. Similarly, the three squares on the right and left three are located using the two alar and saddle landmarks.

The nine square windows centred on the nine landmarks are shown in Figure 5.7(a). The edge of each window can be resampled to a series of points and the local features are found by calculating the differences between those edge points and central point in each window, which is an extension of the proposed local shape descriptor presented in 5.3.1 using the following equation:

$$\begin{cases} F_{Depth} = \sum_j^n \sum_i^m (E_{Depth}(i) - C_{Depth}(j)) \\ F_X = \sum_j^n \sum_i^m (E_X(i) - C_X(j)) \\ F_Y = \sum_j^n \sum_i^m (E_Y(i) - C_Y(j)) \\ F_Z = \sum_j^n \sum_i^m (E_Z(i) - C_Z(j)) \end{cases} \quad (1)$$

where n is the number of windows, m is the number of points on the edge, E refers to the value of each edge points and C denotes the central points. In this experiment, each of the nine windows has 16 edge points resulting in feature sets of size 9×16 = 144 for features extracted from each component. F_{Depth} is depth component and F_X, F_Y and F_Z are SN_x, SN_y and SN_z components.



(a) (b)

Figure 5.7: 9 landmarks and their corresponding patches on the nasal region

5.4.3 Recognition Performance Evaluation Using Different Scales

The influence of the scale of the window on the recognition performance can be evaluated by altering the vertical distance from centre to edge of the square patch. To this end, sizes ranging from 4 to 26 are used to evaluate the recognition performance using the Bosphorus database under two identification scenarios: the KFA classifier using cosine distance and the leave-one-out (LOO) classifier using the city block distance.

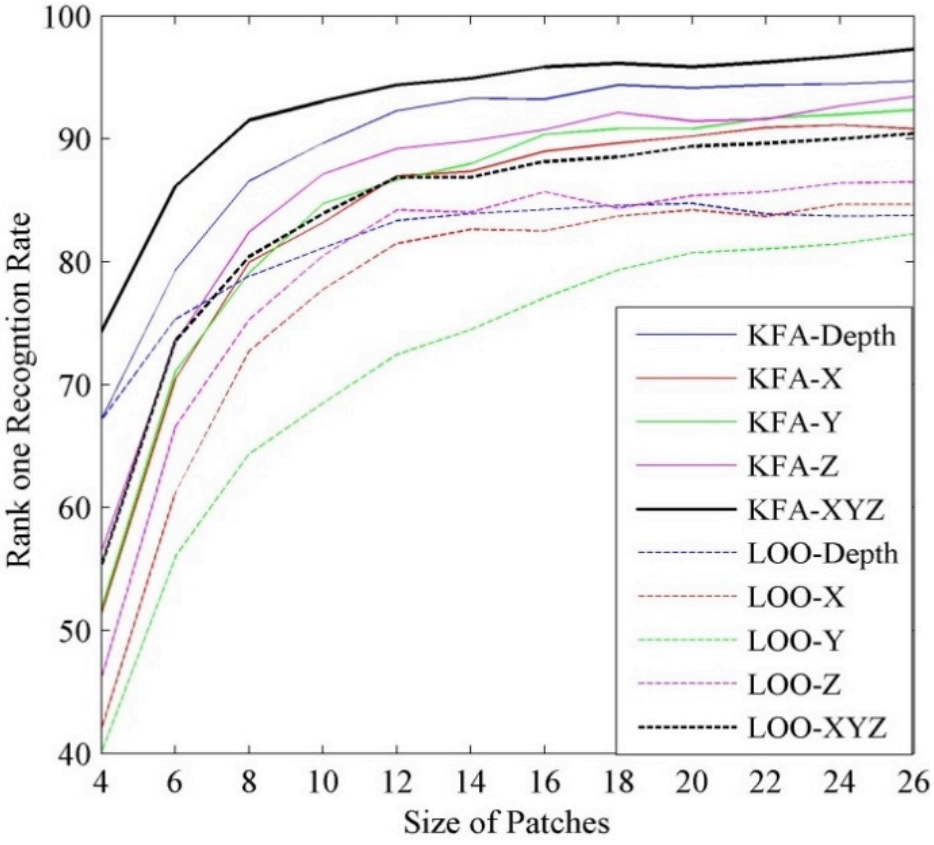


Figure 5.8: Rank one recognition rate calculated by both KFA and LOO classifier using different size of windows (patches) on the Bosphorus database.

The LOO scenario employs every capture as a probe against a gallery of all other captures from different identities. The benefit of using LOO is that all sessions are used for the probe and gallery captures, which makes it a far more efficient scenario than using a fixed partition. The L1-norm [74], city block distance is used to compare the probes and gallery captures as it has been shown to have better discriminatory power when the feature space is sparse.

The R_1RRs for the KFA and LOO classifiers using different patch sizes are shown in Figure 5.8. Although the R_1RR improves with increasing scale for all components, it remains stable with only a small improvement after the patch size reaches 12. Thus 12 is a good compromise scale, as it covers the whole nasal and adjoining regions without the risk of potential occlusions that affects larger patch sizes.

5.5 Conclusions

In this chapter, the use of nasal and adjoining cheek regions have been explored as an expression robust biometric for 3D face recognition. Focusing on these regions is advantageous as they have been shown to be more invariant to variations due to universal expressions and natural occlusions. Thus, extracting features from these stable patches on the nasal region and its environs has a high potential to build a more powerful 3D face recognition algorithm.

In addition, a novel 3D face local shape representation method is proposed to provide a simple way to generate a small feature set that produces high recognition performance. A set of features from the 3D face are found by calculating the depth differences between central points and eight neighbouring points from horizontal curves on the face surface. Experiments show that using the depth differences from three horizontal curves as a feature vector produces a better recognition performance than using the depth values on the curves directly, providing motivation for further work. More localized depth and normal differences were found by including additional points from above and below each horizontal curve and experiments for both the identification and verification scenarios demonstrated that a good performance can be achieved with a relatively small feature set.

Using the proposed local shape descriptor, a further analysis on the discriminatory power of 24 patches on the nose and adjoining cheek regions was evaluated, producing a set of novel discriminative maps for both depth and surface normals (x, y and z). On the basis of the resulting discriminative maps, an extension of the new local shape descriptor was applied to

the nine patches on the nasal and adjoining cheek regions. Results showed that the work has potential for use as an expression robust 3D face biometric or possibly a pattern rejector.

A series of experiments have been presented to evaluate the recognition performance of the proposed multi-component and multi-scale local shape descriptor on the nasal and adjoining cheek regions, focusing on addressing the problem of expression variations on the human face. This method has not evaluated the single training sample scenario. Therefore, one area of future work will focus on exploiting an effective subset from the original features using some feature selection methods or fusing other features to build a new feature vector with more discriminative information for matching.

Chapter 6 3D Face Recognition Using Reconstructed Captures from Photometric Stereo

6.1 Introduction

In the past decades, many 3D data acquisition methods have been developed to produce a better approach to capturing the 3D human face and accommodate both the research and commercial requirements. An appropriate 3D acquisition approach can provide geometric information of the facial surface and hence potentially address the problems of illumination, expression, pose and occlusion variations. The acquisition system should also effectively and accurately capture the human face within a short time period and require minimal cooperation from the candidates. The captures used in most of the literature are from laser scanners or structured light which generate a higher accuracy 3D point cloud. However, these widely used acquisition systems have limitations in the real world applications, for example the requirement of candidates' cooperation. The Photoface device was proposed to address this problem which aims at providing a low-cost and unconstraint 3D face acquisition system [9, 17].

Feature space manipulation on various types of facial captures are very challenging in biometrics as the resulting captures from different acquisition systems contain varying characteristics or the extracted features might be database specific. For example, the Photoface captures have higher resolution surface normals but more distorted depth information in comparison with other widely used captures. Very few works have investigated the use of the Photoface captures for face recognition, even though they possess great potential to produce promising recognition performance in real world applications. Therefore, this chapter addresses the problem of extracting discriminative features from the Photoface captures for face recognition.

Inspired by the successful use of nose recognition algorithm proposed in [7] and the explorations of Zafeiriou *et al.* [9] and Moorhouse *et al*. [3] on the Photoface captures, this chapter investigates the recognition performance by matching curves on various components

(depth, surface normals and shape index) captured or reconstructed from photometric stereo. As the depth information can be successfully aligned by PCA, instead of the original SNs the SNs calculated from the depth is applied to the whole chapter for comparison. The main contributions are as follows:

- Development of preprocessing, alignment and landmarking algorithms for the reconstructed captures
- Discriminative features extraction from the shape index map.
- Multi-component feature space manipulation (depth, surface normals, shape index) and feature-level fusion to build a more efficient feature set.
- A comparison of recognition performance among the Photoface, FRGC and Bosphorus databases.
- 'One training sample' scenario evaluation on the 3DE-VISIR database

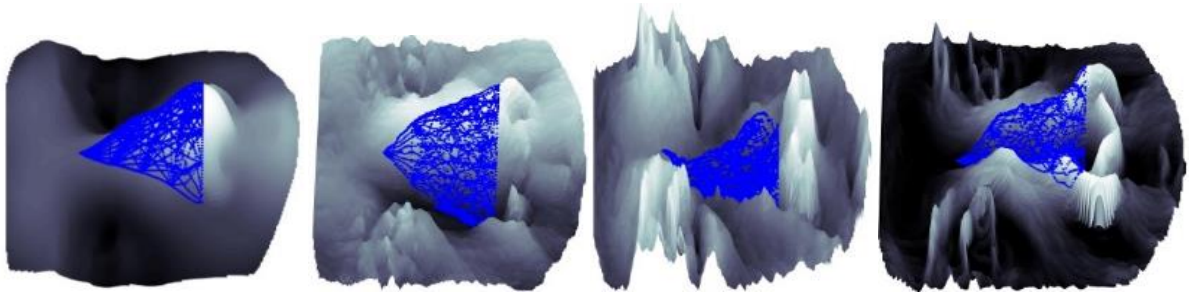
6.2 Landmarks and Curves on the Nasal Region

To define a set of curves, the 1-16 landmarks shown in Figure 4.2 are automatically detected on the perimeter of a triangular structure containing the nasal region [7]. In Chapter 4 and [138], after the nasal landmarks localization, the cheek landmarks were found by calculating two types of distance from the nose tip: a proportional distance and a constant distance. The main difference between these two methods is that the accuracy of the proportional distance is based on the location of previous detected alar grooves. The results obtained in Chapter 4 demonstrate that the constant distance outperforms the proportional one as the alar grooves are less consistent than the nose tip.

Inspired by the benefit of using the constant distance, this chapter proposes to use the constant distance to define the nasal region. The potential benefits of this are the preservation of the within-class similarity and the between-class scatter. The average distances from the detected root to nose tip in the Photoface and 3DE-VISIR databases are defined as Drt_p and Drt_v, respectively, where the subscript p and v denote "Photoface" and "3DE-VISIR", respectively. Also, the average distances, Daa_p and Daa_v, between two alar grooves of two databases are calculated. As shown in Figure 4.2, the Drt_p and Drt_v are the average length of L1L9 and the Daa_p and Daa_v are the average length of L5L9. Using the calculated average coefficients, the nasal root and alar grooves are localized and the nasal region is cropped from the whole face.

6.3 Nose Identification by Matching 75 Curves from Depth and Normals

In this section, 75 nasal curves found on the four components (depth, SN_x, SN_y and SN_z) shown in Figure 6.1 are evaluated under identification scenarios. To make a comparison between different types of captures, 360 captures of 18 subjects are selected from the Photoface, FRGC and Bosphorus databases, respectively.



(a) Depth (b) SN_x (c) SN_y (d) SN_z

Figure 6.1: 75 nasal curves representations on the depth and surface normals components

6.3.1 Recognition Performance for 75 Nasal Curves

In Chapter 4, instead of using 50 points, each curve is resampled to only 15 points and the results demonstrated that using fewer points on the nasal and cheek regions can produce a comparable recognition performance, even though the dimensionality of the feature space is greatly reduced. Therefore, in this chapter, 75 nasal curves with 15 points per curve are extracted for recognition. To make a comparison of the recognition performances generated from different types of captures, the subset (360 captures of 18 subjects) of each database is employed for evaluation.

To reduce the dimensionality and find the most discriminative curves or curves combination, FSFS was applied to the 75 curves, combined with the leave-one-out cross validation and nearest neighbour classifier. The R_1RRs obtained from set of captures for the three databases and different components (depth, SN_x, SN_y and SN_z) are shown in Figure 6.2, which describes the changes of R_1RRs when the number of selected curves varies from 1 to 30. The recognition performances are not provided when more curves are selected as all the R_1RRs reduce for ≥ 30 curves.

The R_1RRs for the FRGC database, shown in red, outperform those for the Bosphorus (blue) and Photoface (black) databases for the reason that the captures from Bosphorus and Photoface databases contain more expression variations. In addition, for all three databases, features

extracted from the surface normals (SN_x, SN_y and SN_z) are shown to be more discriminative than those from the depth component. Only the features extracted from the depth component were explored in [7] and so there is potential to further evaluate the surface normals or combine both depth and surface normals to build a stronger feature set.

The L9L1 curve, from the nasal root to tip, is always first selected (except for the SN_x map) and makes a significant contribution to the overall recognition performance. L9L1 is located on the central profile of human face, which can better and more robustly describe the nasal surface changes. However, in the SN_x map, this curve suffers significant feature-level changes as long as it is not localized accurately, for example the sign of the points on the L9L1. For the depth, SN_y and SN_z components of the FRGC database and the SN_y component of the Photoface database, all the R_1RRs of L9L1 exceed 80%.

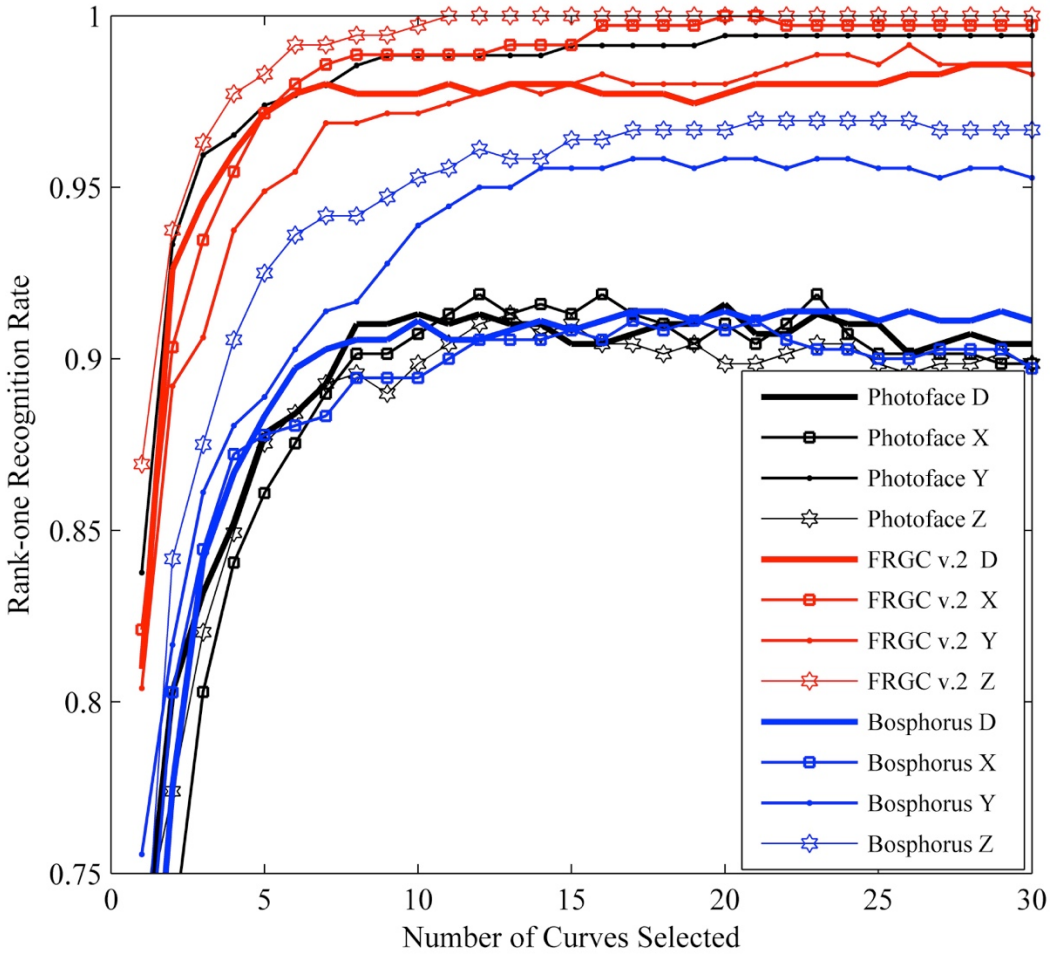


Figure 6.2: R_1RRs against the number of curves selected by the FSFS and leave-one-out. Four components, depth, SN_x, SN_y and SN_z, from each subset (360 captures of 18 subjects) of the Photoface, FRGC and Bosphorus databases are used in this experiment. Only the first 30 selected curves are provided for recognition performance comparison.

6.3.2 Recognition Performance for the Landmarks Localized by the Constant Distance

For the Photoface captures, the nasal root and two alar grooves are not always robustly localized, as few local nasal structures are provided in the reconstructed depth map. Therefore, instead of using the NCM landmarking [7] employed in Section 6.3.1, this section uses two constant distances, Drt_p (from the tip to root) and Daa_p (from the tip to alar grooves), to localize the root and alar grooves, resulting in 75 curves on a fixed sized nasal region. For Daa_p, the nose tip is located in the centre, which results in an isosceles triangle covering the whole nasal region. This means that whatever is the real size of nasal region, all the captures use the same window to crop the nasal region.

The R_1RRs shown in Figure 6.3 are calculated by the same settings, as for Figure 6.2. This enables a direct comparison with the recognition performances by using the NCM landmarking for the Photoface database from Figure 6.3. It is clear that the features extracted from the fixed sized nasal region described by green curves outperform the region found by NCM landmarking, which proves the conclusion stated in [138]. The constant distance is likely to preserve the within-class similarity and between-class scatter, as the captures from the same subject contain similar geometric distributions after pose calibration.

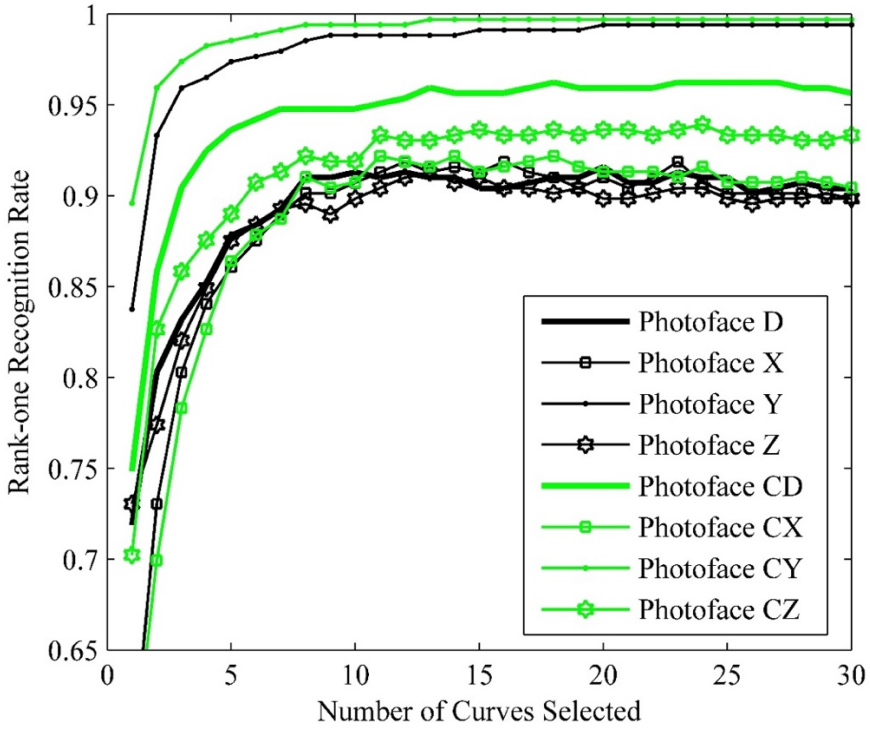


Figure 6.3: A comparison of R_1RRs between fixed sized (constant distance) and calculated (NCM landmarking) nasal region in the Photoface database using depth (D), SN_x (X), SN_y (Y) and SN_z (Z) components. Only the first 30 selected curves are provided for comparison.

6.3.3 Classification Evaluation Using the Features Extracted from the SN_y Map

Compared with the other two databases, the recognition performance of the features extracted from the SN_y map in the Photoface database is apparently better than those from the depth, SN_x and SN_z maps. Using only 9 selected curves on the SN_y map, the R_1RR can reach 98.84%, which demonstrates the potential of extracting discriminative features for face recognition from the Photoface captures. Compared to the SN_x and SN_z components, these 9 curves on the SN_y map produce a higher recognition performance. Therefore, instead of using the less accurate depth map of the Photoface captures, it is very promising to investigate new discriminative features on the surface normals components.

A further classification evaluation is explained in this section to demonstrate the recognition performance of features extracted from the SN_y map. KFA [139] with the polynomial kernel is applied for nonlinear projection and dimensionality reduction. Instead of using 18 subjects with 20 captures per subject, this evaluation takes all the captures in the Photoface database into account and tests the recognition performance using from 1 to 10 of training samples. Figure 6.4 shows the R_1RRs against the number of training samples per subject.

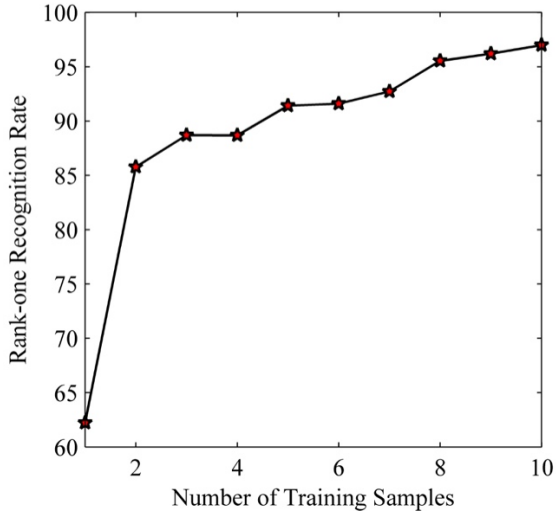


Figure 6.4: R_1RRs against the number of training samples using the selected curves from the SN_y map of the Photoface captures. All the captures in the Photoface database are used in this evaluation.

When the training set only contains one or two samples for each subject, the R_1RRs (62.22% and 85.79%, respectively) are very comparable to the results reported in [9]. Furthermore, the efficiency is much improved using the proposed method as only the human nasal region and a small number of features (135 values) are considered for face recognition. When the number of training samples increases to 10, the R_1RR reaches 97%, demonstrating that the recognition performance can be greatly improved provided there are enough training captures of each

identity. Instead of exploiting other feature selection and classification strategies, there is great potential for improved performance by fusing with the features extracted from other components.

6.4 Using the Shape Index for Nose Recognition

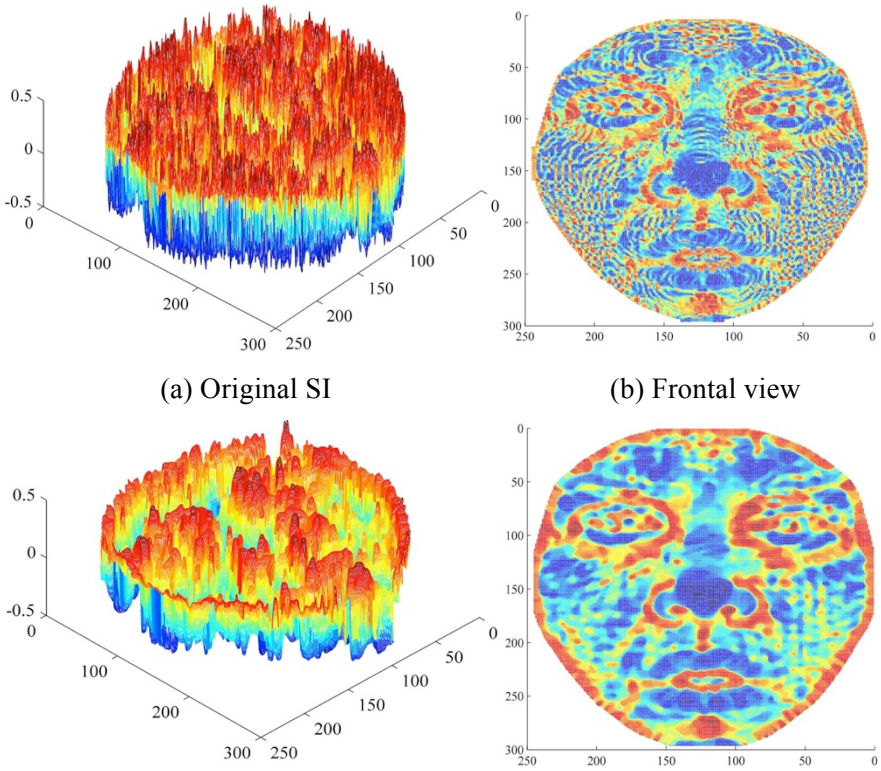
Although many published works have considered the SI as a new type of 3D face representation to generate discriminative features, it is still an under-explored representation [18]. Huang *et al*. [142] used a multi-scale local binary pattern depth map together with the SI map to increase the distinctiveness of smooth range faces. Vijayan *et al*. [132] also explored the use of the SI map for 3D face recognition of twins. In [19], a comparison of a series of 3D facial features showed that the SI map outperformed the depth map but was worse than the point cloud and surface normals.

Despite some advantages, using the original SI map may encounter some unexpected problems. For example, the SI coefficients calculated on the planar region are relatively noisy as the curvatures of such regions are low, which makes it hard to extract robust features on such regions. Therefore, the original SI map is not an appropriate representation for describing all the face, especially for the cheeks. To avoid this drawback, the SI features can be restricted to the less flat regions, for example on the nasal region.

To smooth the original SI map, some effective filters that have successfully been applied to image denoising can provide a good solution to address this problem, for example the median filter. In addition, the curvedness [35], a positive number that specifies the amount of curvature, can remove the noisy data by thresholding the original SI map. In the next section, the discriminative SI features are extracted from the nasal region and the recognition performances obtained by using different SI smoothing approaches are also discussed.

6.4.1 Median Filtering the SI Map

Figure 6.5(a) and (b) demonstrate the original SI map of one sample capture selected from the Bosphorus database. The noisy points, shown in both the frontal and side views, might affect the region segmentation and feature extraction. A recent summary and evaluation of different denoising methods applied to 3D face recognition suggests that the median filter is a good choice for denoising [135]. Therefore, the median filter is employed to smooth the original SI map and the resulting denoised SI map is shown in Figure 6.5(c) and (d). Experiments have shown that 3×3 is a suitable size of median filters for denoising both the depth and SI maps.



(a) Original SI (b) Frontal view

(c) Denoised SI (d) Frontal view (denoised)

Figure 6.5: Median filter applied to the original SI map calculated from one sample capture selected from the Bosphorus database.

The set of nasal curves used in the previous section is also extracted from both original and denoised SI maps and the corresponding recognition performances are shown in Figure 6.6 (blue curves). The R_1RR is much higher when the median filter is applied and, in particular, the R_1RR of the curve L9L1 increases by ~17% after denoising. To compare the recognition performance of features extracted from both original and denoised SI maps among different databases, a subset of the FRGC v.2, Bosphorus and Photoface databases are used for evaluation, which employs the same number of captures and experimental settings (FSFS based feature selection of 75 nasal curves) as the comparisons described in Section 6.3.

In general, as can be seen from Figure 6.6, the recognition performance of the Photoface captures, shown by black curves, outperforms the Bosphorus (blue) and FRGC v.2 (red) databases. The R_1RRs of the curve L9L1 experience ~20% improvement for all three databases. However, for the combination of curves, using the denoised SI map brings more recognition performance improvement in the FRGC v.2 and Bosphorus databases, which is about 5-10% higher than using the original one. In contrast, the median filter brings less improvement when more than 4 curves are selected in the Photoface database. Denoising is of limited benefit as the captures in the Photoface database are relatively smoothed.

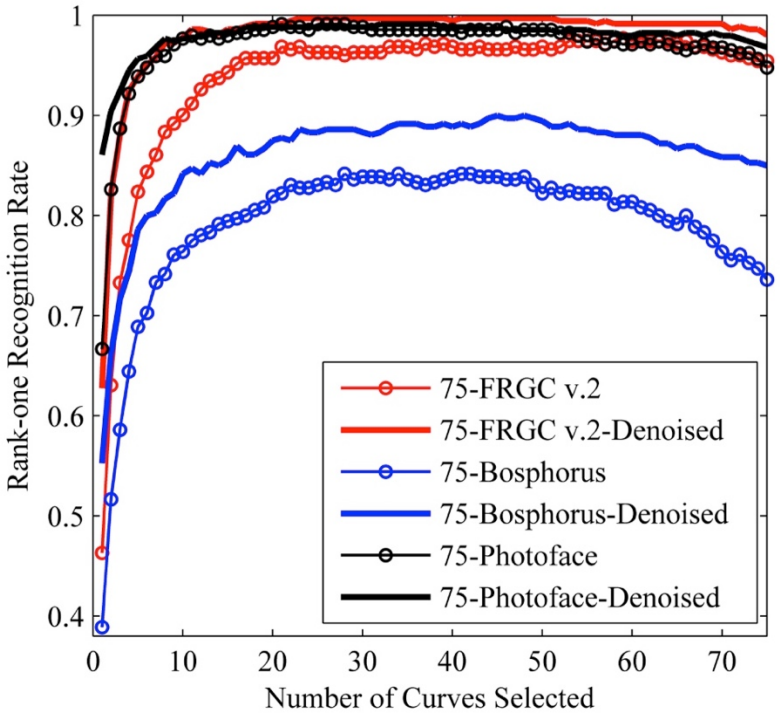
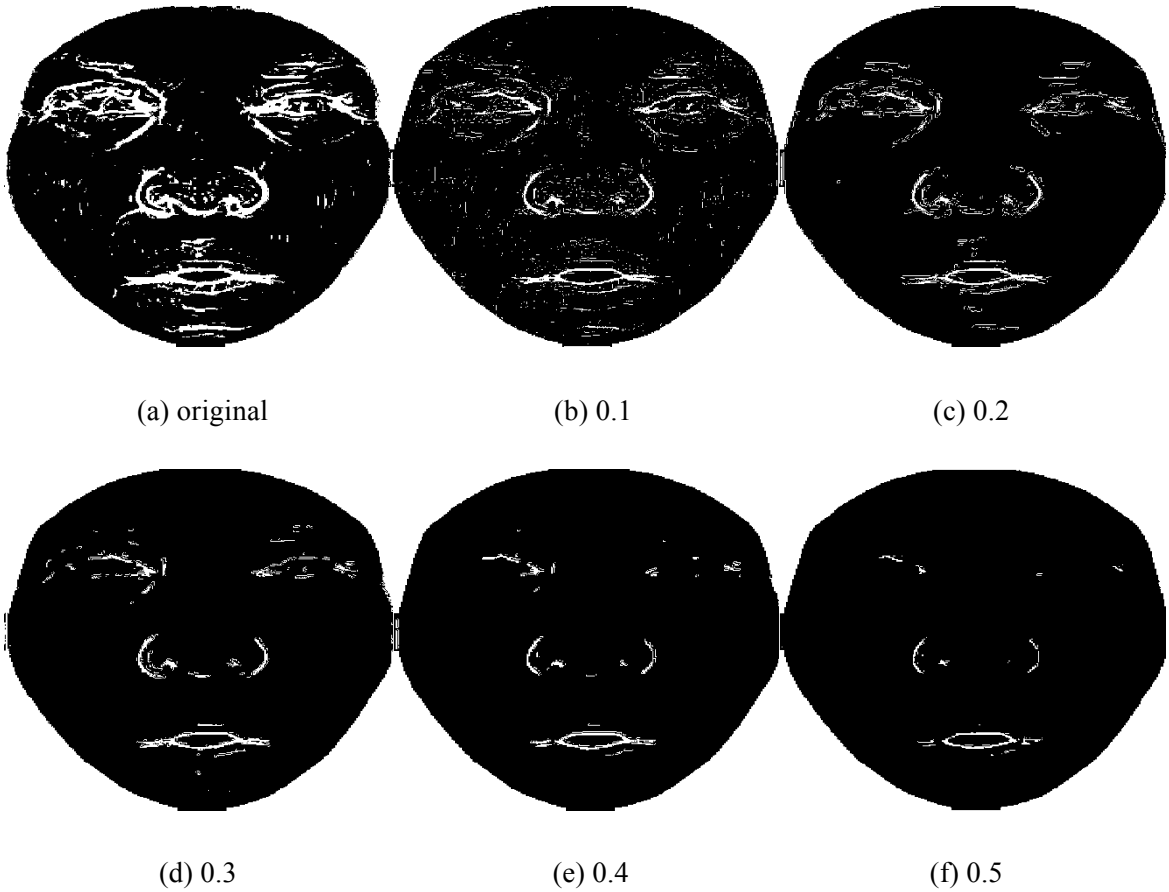


Figure 6.6: A comparison of the recognition performance of the 75 curves extracted from both original and denoised SI maps among three databases. The curves are sequentially selected by FSFS.

6.4.2 Curvedness Thresholding

The curvedness can be considered as a rotation invariant gradient operator, which measures the degree of regional curvature. It has true geometrical significance as it is coordinate independent. Therefore, curvedness has the potential to be considered as a new and effective representation of the 3D facial surface to describe curve degree of each point.

In Figure 6.7, the original capture from the Bosphorus database is thresholded by different curvedness values. With the thresholds increasing, the regions contain a high degree of curved surfaces are more prominent, for example the nasal alar grooves and mouth. To remove the noise points on the relatively flat regions in the SI map, the threshold 0.1 is first applied. Each point in the SI map is first smoothed by the median filter and used for feature extraction if the curvedness is bigger than 0.1. However, as can be seen from Figure 6.8, using the curvedness thresholding did not improve the recognition performance. Data loss might be the main reason for this phenomenon as some useful features are removed during the thresholding phase. Therefore, although curvedness thresholding can provide a good approach to smooth the SI map, some discriminative information might be lost and selecting an appropriate threshold for face recognition is still a challenging issue. In comparison, the median filter or other effective filters are more suitable for low complexity SI map denoising.



(a) original (b) 0.1 (c) 0.2

(d) 0.3 (e) 0.4 (f) 0.5

Figure 6.7: An example of the curvedness thresholding by different parameters using the capture selected from the Bosphorus database

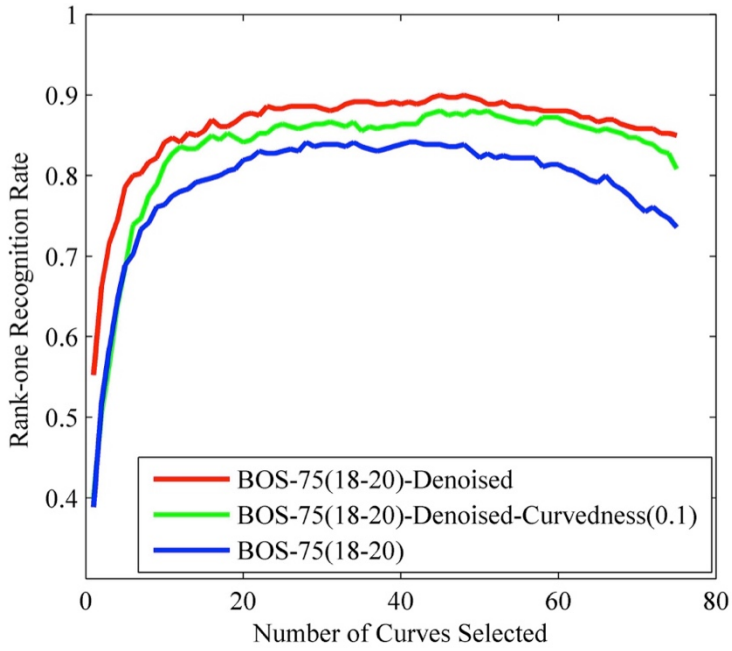


Figure 6.8: R_1RRs against the number of curves selected by FSFS using the capture selected from the Bosphorus database. It is a comparison of the recognition performance using the SI maps denoised by the median filter and curvedness thresholding.

6.4.3 Gabor Filtering the SI Map

The Gabor filter, a Gaussian kernel function modulated by a sinusoidal plane wave, is one of widely used image denoising filters and has been introduced to smooth the depth map for robust feature extraction [143]. The recognition performances after introducing the Gabor filters with different parameters on the SI map are shown in Figure 6.9 and Figure 6.10.

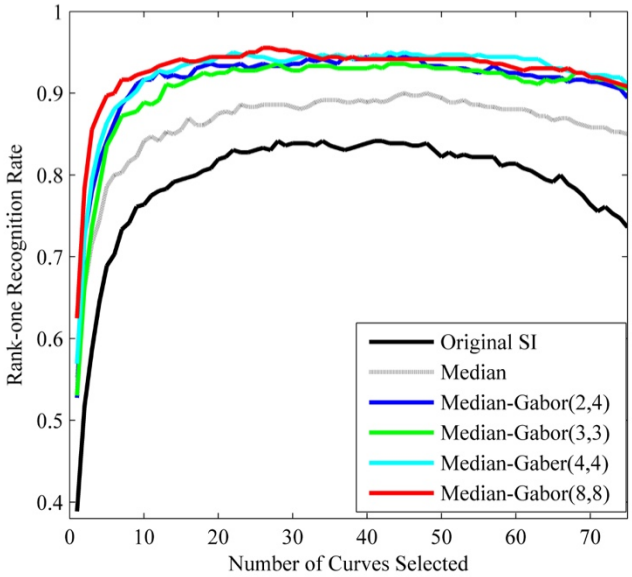


Figure 6.9: R_1RRs against the number of nasal curves selected for face recognition using the Gabor filters with different variances. For example, in Gabor (2,4), "2" and "4" are the variances along *x* and *y* axis, respectively. The frequency is set to "16". 360 captures of 20 subjects from the Bosphorus database are used.

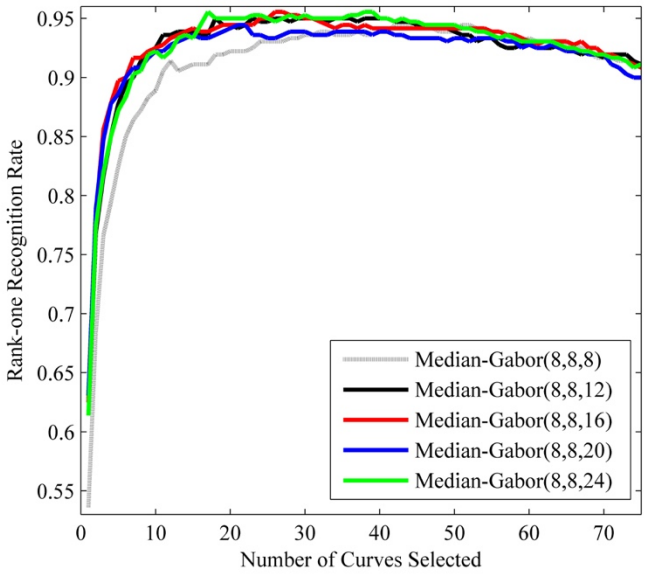


Figure 6.10: R_1RRs against the number of nasal curves selected for face recognition using the Gabor filters with different frequencies. For example, in Gabor (8,8,16), "8" and "8" are the variances and "16" is the frequency. 360 captures of 20 subjects from the Bosphorus database are used.

The denoised SI map, $G(x, y)$, can be obtained by

$$G(x,y) = \frac{1}{2\pi S_x S_y} * e^{\frac{-1}{2}[(\frac{x}{S_x})^2+(\frac{y}{S_y})^2]} * \cos[2\pi f\sqrt{x^2+y^2}] \quad (1)$$

where S_x and S_y are the variances along x and y axes, respectively. f is the frequency of the sinusoidal function. As can be seen from Figure 6.9, the R_1RRs have been significantly improved when the Gabor filters were applied to the denoised SI map by the median filter. The Gabor filters with different variances and frequencies are explored in Figure 6.9 and Figure 6.10, respectively, which demonstrate similar contribution to the recognition performances improvement.

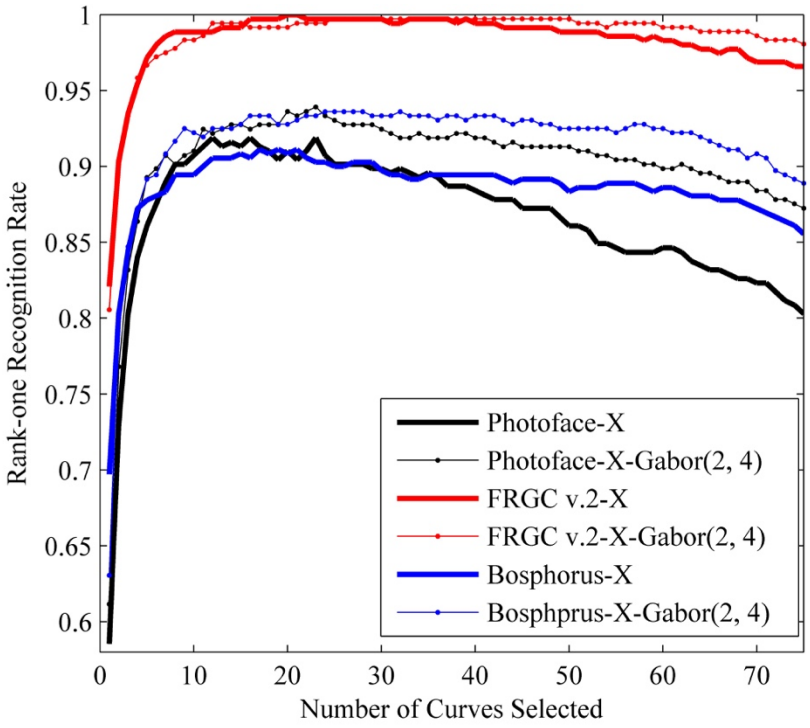
6.5 Denoising of the Surface Normal Components

Denoising is always executed in the preprocessing step to smooth the original facial captures for improving the accuracy of feature detection. In this section, to extract more robust features from the surface normals maps or enrich the diversities in the feature space, the Gabor filters used in the SI map are further applied to the data smoothing before feature extraction. Features are extracted from different resulting maps filtered by various Gabor wavelets as some distinct features or structures can be found in specific frequencies or scales.

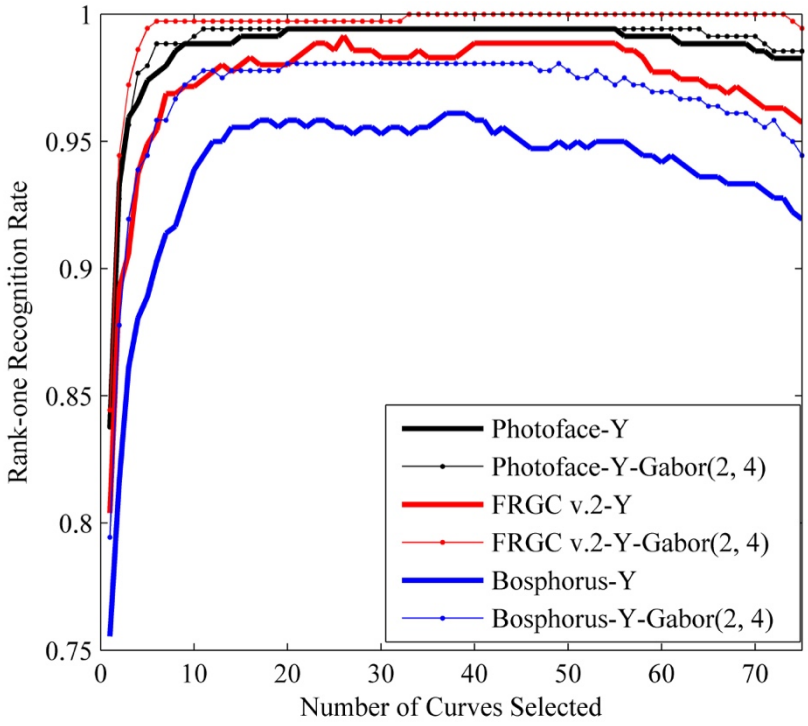
For all three databases, the surface normals are calculated from the depth information, which might still contain some noisy points even though the depth map has been already smoothed by the median filter. Any residual noise in the depth data can be amplified during the estimation of the surface normals. Therefore, it is interesting to investigate the denoising effect on the surface normals computed from various types of captures. The recognition performances of 75 nasal curves on the surface normals maps from three databases are evaluated in this section, using the same FSFS based feature selection method and leave-one-out cross validation, as described in Section 6.3.

As shown in Figure 6.11(a)-(c), the denoised surface normal components provide different contributions to the recognition performance improvement in the three databases. In Figure 6.11(a), for 75 nasal curves found from the SN_x map, there is a 2-3% improvement on the Photoface and Bosphorus databases when 20 curves selected. The recognition performance increases by ~8% when all the 75 nasal curves are used in the Photoface database. In Figure 6.11(b), the captures from the FRGC v.2 and Bosphorus databases experience 2-3% increase using the denoised SN_y map and there is no significant improvement in the Photoface database.

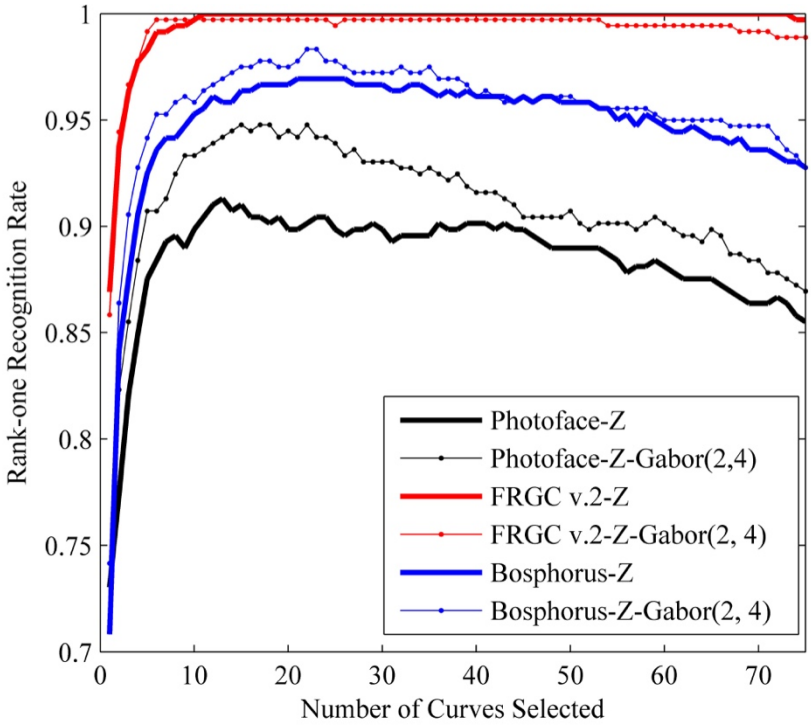
In contrast, as shown in Figure 6.11(c), the improvement of using denoised SN_z map in the Photoface database is prominent, about 5%.



(a) denoising applied to SN_x



(b) denoising applied to SN_y



(c) denoising applied to SN_z

Figure 6.11: R_1RRs against the number of nasal curves selected for face recognition using the Gabor filters. A comparison of the recognition performances of original and denoised surface normals components of 360 captures of 20 subjects from the FRGC v.2, Bosphorus and Photoface databases are used. The SNs used in the Photoface are calculated from the depth map.

These results confirm the conclusions found in Section 6.3 and 6.4, which indicate that 3D captures obtained from different types of acquisition systems might contain various characteristics for face recognition. Different feature space manipulation strategies for each type of data are able to contribute to the recognition performance. When 20 denoised nasal curves are selected, the improvements of the R_1RRs in three databases are further examined in Table 6.1. The results show that further data smoothing on the SN_x and SN_z components before feature extraction is necessary for the Photoface database and the effect of denoising is actually modest for higher accuracy captures.

Table 6.1: The improvements of R_1RRs when 20 curves selected from each component

	SN_x	SN_y	SN_z
FRGC v.2	——	~2%	——
Bosphorus	~3%	~3%	~1.5%
Photoface	~2%	——	~5%

6.6 Classification Performance Evaluation on the 3DE-VISIR Database

The 3DE-VISIR database was originally proposed for expression classification and contains at least three sessions, positive, negative and neutral, for each identity [18]. The captures are obtained from both visible (VIS) and near infrared (IR) lights. 3D facial information estimation using these two types of illumination produces different influence on the reconstruction, resulting in 3D captures with different characteristics.

As a subset of the 3DE-VISIR database, the IR captures possess a lot of advantages for 3D face recognition and can significantly contribute to the real world applications. Hansen *et al*. [10, 18] proposed the use of near infrared light to capture the 3D face as it is very directional and gives a more accurate overall depth reconstruction than visible light. Furthermore, the acquisition systems using the near infrared light are more covert and less intrusive in real world applications. However, some fine surface details might be lost due to sub-surface scattering [144], even though the overall reconstruction by using the IR light is better than the VIS light [18].

Compared to other face databases containing expressions, the 3DE-VISIR database does not contain multiple expressions for each subject and only categorizes the expressions into positive, neutral and negative groups. Such classification provides a wider range of happy and sad related expressions and makes candidates feel more relaxed, which is likely to result in more natural expressions [18]. Therefore, it is a useful source for expression robust features evaluation.

In addition, the 'one training sample' scenario is another challenging issue in biometrics, which requires that the features extracted and selected for each identity should be more effective and discriminative for matching. It is interesting to apply the nasal curves found from the Photoface database to this new database. Therefore, in this section, the neutral capture of each identity is used for training and the other two captures from the same identity with positive or negative expressions are used for testing. To make a comparison of the VIS and IR captures, the recognition performance of 'one training sample' scenario using both positive and negative captures from VIS and IR groups will be evaluated individually.

6.6.1 Feature Extraction and Recognition Performance Evaluation Using VIS Captures

The VIS part in the 3DE-VISIR database can be considered as an extension of the Photoface

database as their acquisition conditions use the same setting. As a consequence, features extracted by 75 nasal curves and FSFS based feature selection used in the Photoface database can be directly applied to this database.

After drawing the 75 nasal curves, feature selection is applied to each component, which results in different curves combinations and can build a stronger classifier. The resulting R_1RRs of each component tested on the VIS part are illustrated in Table 6.2. Features extracted from the depth, SN_x, SN_z and SI components produce similar recognition performance no matter whether the positive or negative probes are used. The R_1RRs generated from the surface normals and SI components significantly outperform the depth component. Furthermore, the resulting R_1RRs are higher than those reported in [9].

Table 6.2: R_1RRs of the best curves combinations of four components using VIS captures under 'neutral vs. non-neutral' one training sample scenario.

	R_1RR	
	Positive	**Negative**
Depth	53.49%	51.72%
SN_x	70.93%	70.11%
SN_y	80.23%	72.09%
SN_z	75.58%	71.26%
SI	73.26%	71.26%

Using the positive captures, the R_1RR of SN_y clearly outperforms the other four components and the R_1RR of 80.23% is produced from only 10 curves from SN_y map. The feature-level fusion of the SN_y and SN_z components produces a R_1RR of 87.21%, which is very competitive in 'one training sample' scenario as the size of features set is very small and the nasal curves extraction is relatively low complexity. For the negative captures, the R_1RR (72.09%) of SN_y component is also the highest but is much lower than that using the positive probe. The curves used for matching in each component and groups (positive or negative) are varying as the FSFS results produce different numbers or combinations of curves in each feature selection.

6.6.2 Feature Extraction and Recognition Performance Evaluation Using IR Captures

Compared to the landmarking on the VIS captures, the nasal root localization fails on nearly all the IR captures and the consistency of two alar grooves is less accurate than using VIS captures. As can be seen from Figure 1.1, there are not enough fine details provided on the

depth map of IR captures, especially for the nasal and adjoining regions, although the overall estimation of depth information is more accurate [18]. As a consequence, it is hard to apply the well-designed landmarking approaches proposed for the captures with higher accuracy depth map directly to the IR captures. Alternatively, the method proposed in Section 6.3 provides an effective solution to address this problem, which used the constant distances to localize the root and alar grooves.

Using the landmarks found by the constant distances, 75 nasal curves extracted from depth, surface normals and SI maps are used to investigate the IR captures and the resulting R_1RRs are illustrated in Table 6.3. The recognition performances of each component tested on the IR captures outperform those of the VIS captures. Using positive captures as the probe, R_1RRs have ~10% improvements for SN_x and SN_z components and ~5% improvements for the depth and SI components.

Table 6.3: R_1RRs of the best curves combinations of four components using IR captures under 'neutral vs. non-neutral' one training sample scenario. A fixed sized structure is applied instead of using the detected landmarks.

	R_1RR	
	Positive	**Negative**
Depth	56.98%	60.92%
SN_x	80.23%	78.16%
SN_y	79.07%	77.01%
SN_z	80.23%	80.46%
SI	79.07%	78.16%

In contrast, for the negative probes, the R_1RRs of all the components increase by ~10%. Moreover, the curves selected from the SN_y map do not have the best performance and the surface normals components are still better than the depth and SI. Compared to using the VIS captures, the IR captures have higher probability to produce discriminative features for recognition, which mainly results from the high accuracy of surface normals reconstruction using IR light.

6.7 Discussion and Conclusion

In the past decade, many works have focused on developing recognition algorithms for captures with higher accuracy depth maps obtained by the laser scanners or structured light. However,

these 3D data acquisition systems that can generate captures with higher accuracy are hard to employ in the real world applications. The Photoface device was developed to address this problem and brought the potential to using this new type of captures for 3D face recognition. Therefore, to boost the development of 3D face recognition system, this chapter has an emphasis on employing existing well-designed algorithms for discriminative features extraction on the Photoface captures. To make a comparison of the recognition performances of various types of 3D captures, 360 captures of 20 subjects are selected from the Bosphorus, FRGC and Photoface databases, respectively. Using the surface normals components of the Photoface captures outperforms those of the Bosphorus and FRGC databases. The use of the shape index map and the recognition performances of the denoised shape index and surface normals are also explored in this chapter.

To further extract the expression robust features for 3D face recognition, a subset of the Photoface database, the 3DE-VISIR database, is used to evaluate the recognition performance. The results demonstrate that it is likely to extract more robust features from the SN_y and shape index maps in the VIS part. The recognition performance can be significantly improved using the captures reconstructed from IR light. However, the landmarking approaches proposed for captures with higher accuracy depth maps are not so effective for the IR captures as some protrusions, such as the nose and eye sockets, on the depth maps reconstructed from surface normals are flatter than their real-world counterparts, mainly resulting from the data loss during the reconstruction. Therefore, novel landmarking algorithms designed for such kind of 3D captures should be further explored. The original surface normals calculated directly from the outputs of Photoface device possess higher resolution, which will be further evaluated for the landmarking of Photoface captures in the next chapter.

A comparative test using other published algorithms is not presented in this chapter as it is the first work that emphasizes the extraction of discriminative features on the nasal region using the Photoface captures. The 3DE-VISIR database was originally proposed for expression classification and therefore for one training sample scenario classification evaluation on the two types of light are not evaluated in other literature.

Chapter 7 Landmarking Using Surface Normals

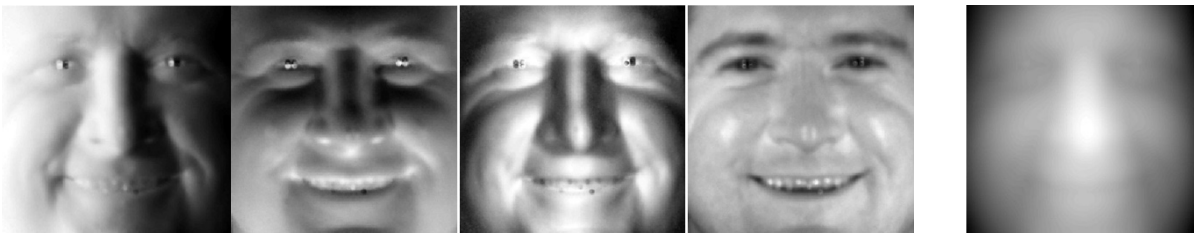
7.1 Introduction

In the previous chapters, the advantages and motivations of using the Photoface captures have been discussed, which demonstrate the high potential to produce good recognition performance. However, as shown in Figure 7.1(f), the depth maps produced from the Photoface surface normals are flatter than their real-world counterparts with most protrusions understated [18], such as the nose tip, eye sockets and the sides of the nose, resulting in failed landmarking on the Photoface database using the curvatures based algorithms. Therefore, simply applying well-designed landmarking algorithms to the Photoface captures is not guaranteed to produce a good performance. New landmarking algorithms designed for such kind of 3D captures should be further explored to address this problem.

Surface normals (SNs) are an effective 3D shape representation and provide a good solution to address this problem, as the original data computed from the photometric stereo captures is full field SNs and albedo. Instead of using depth maps, the SNs of each point determine the orientation of the facial surface and provide a comprehensive description of facial local details and more facial geometrical information. Both the surface normals estimated from the high accuracy depth maps and recovered directly by the Photoface device have been reported to produce better recognition performance than the depth [9, 19], which demonstrate that SNs are a good 3D shape representation with high discriminatory power for both types of 3D captures. It is therefore interesting to explore the use of SNs for landmarking. To provide a good compromise between the accuracy and efficiency, a novel landmarking algorithm is proposed in this work for the captures with higher or lower accuracy of depth maps.

In addition, to improve the overall consistency of the detected landmarks, the proposed approach can address the pose variations in yaw and roll directions and the deviation of nose tip. The Bosphorus and FRGC databases contain many 3D captures with high quality depth maps derived from their 3D point clouds and various expressions. The 3DVISIR database [18] is a small subset of the Photoface database captured by both visible light and near infrared light based photometric stereo acquisition system. These three databases that contain varying types of captures are used for landmarking performance evaluation, which shows the robustness of the proposed algorithm as well as the contribution of recognition performance improvement.

(a) 4 images captured using different light directions



(b) SN_x (c) SN_y (d) SN_z (e) Albedo (f) Depth

Figure 7.1: The outputs and reconstructed 3D captures of the Photoface device using near infrared light. (a) shows four 2D images obtained directly from the acquisition system. Captures from (b) to (e) are the SNs and albedo, respectively, estimated by 4 images. The depth in (f) is reconstructed from the SNs.

7.2 Landmarking on the 3D Captures with High Accuracy Depth Maps

Following [7], the nose tip is first detected by thresholding the SI and then the initial nasal region is cropped by the tip. After Principal Component Analysis (PCA) based alignment, most pose variations are corrected and the tip is further localized. However, some rotations still remain, especially in the yaw and roll directions, due to the unexpected facial occlusions or special facial structures, which are hard to correct using PCA based alignment. Therefore, in this section a novel landmarking algorithm is proposed, which aims to address the remaining pose variations using the nasal geometrical structure and the detected nasal landmarks, as well as further improve the accuracy and consistency of nasal landmarks.

A brief description of the main steps of the proposed approach are shown in Figure 7.2. Using the thresholded SNs maps, the nasal bridge is first localized to approximately correct the roll variations. A new nose tip detection method is proposed, which uses the nasal bridge and nose profile signature. On the basis of the tip, candidate alar grooves regions are found by thresholding the SNs, which can help calibrate self-yaw rotations, relocate the tip and provide a reference map for the final alar grooves localization on the depth map. The nasal root and subnasal are localized in a similar manner to that used for the tip detection. As the nasal root detection is sensitive to roll rotation, a fine roll rotation calibration is proposed and the rotation angle is estimated using the detected root and tip. Finally, five nasal landmarks (the tip, root, subnasal and alar grooves) are localized on the well aligned 3D nose, as shown in Figure 7.2.

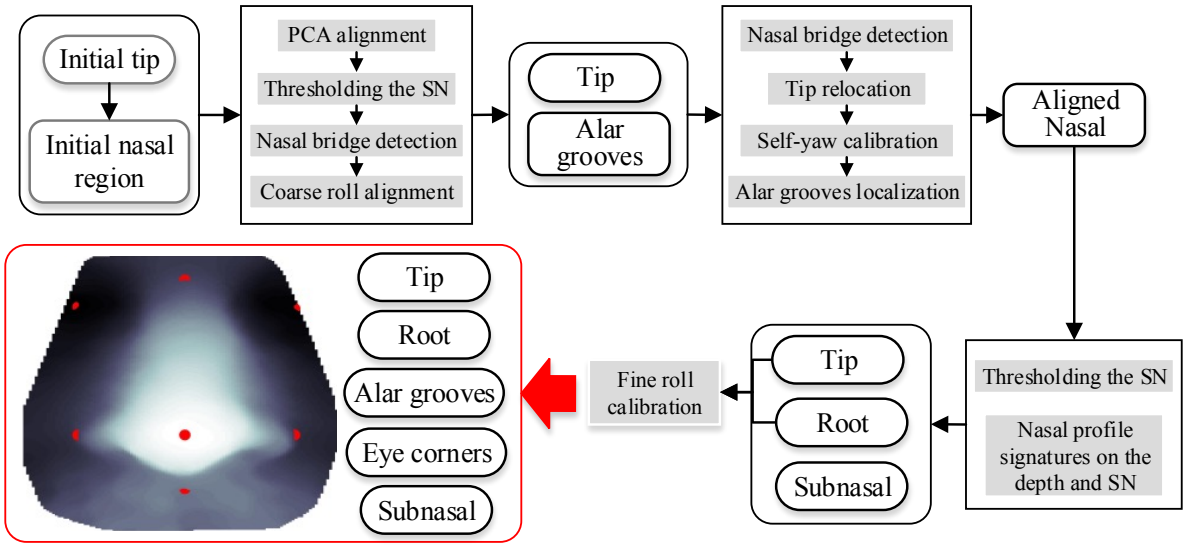


Figure 7.2: The flow chart of the proposed landmarking algorithm: Using the direct outputs of PCA based alignment, the cropped nasal region and the initial tip, the thresholded SNs and nasal bridge are first found for the coarse roll and self-yaw calibration. Then, the location of the root and tip are used for further roll calibration. Finally, the nose tip, root, two alar grooves and subnasal are localized on the aligned nasal region. Eye corners can be localized by thresholding the SNs.

7.2.1 Nose Tip Detection and Coarse Roll Calibration

As shown in the example capture in Figure 7.3(a), some pose variations still remain on the face after the PCA based alignment. These are mainly in the roll direction due to the occlusions by hair, which cannot be addressed by PCA based alignment and might greatly destroy the within-class consistency of the extracted features and the recognition performance. To this end, a coarse roll rotation calibration is first employed to roughly align the nasal region before landmarking. As one of the most salient parts on the nasal region, the nasal bridge is found to estimate the approximate roll rotation angle.

7.2.1.1 Nasal Bridge Detection

As a distinct geometrical structure on the facial surface, the location of the nasal bridge can indicate the degree of roll variations as its centre should be vertical in the frontal view face. Therefore, using this prominent structure can provide a good reference to rapidly correct approximate facial roll variations. Compared to using the depth information, the thresholded SN_x map (TSN_x) provides an effective way to detect the nasal bridge as the SN_x values in this region are close to zero, which can be easily detected by thresholding the SN_x map. The TSN_x region shown in Figure 7.3(c) is a matrix with the same size as SN_x and is found by,

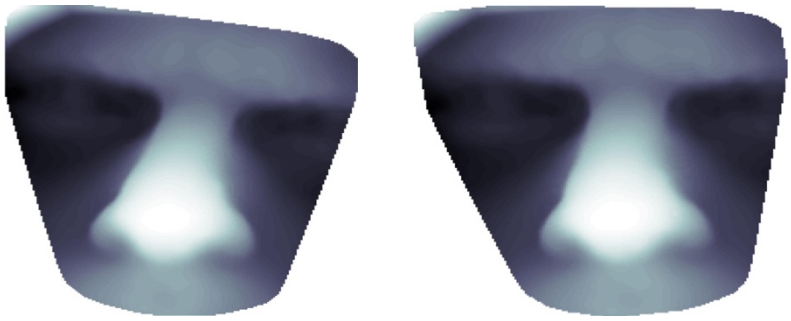
$$TSN_x = \left\{ TSN_{x_{i,j}} \middle| TSN_{x_{i,j}} = f_x(i,j) \right\} \tag{2}$$

where

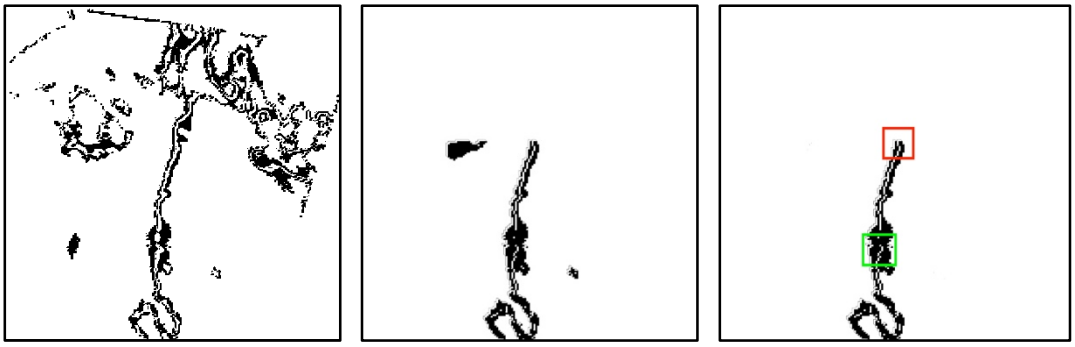
$$f_x(i,j) = \begin{cases} 1 & if\ SN_x(i,j) < T_x \\ 0 & otherwise \end{cases} \tag{3}$$

T_x is set to 0.1 for nasal bridge detection in subsequent experiments.

To detect the nasal bridge in thresholded maps such as that shown in Figure 7.3(c), the main nasal region is first cropped on the basis of the location of the nose tip obtained in the PCA based alignment as shown in Figure 7.3(d). The nasal bridge is located as the largest connected region of the main nasal region as shown in Figure 7.3(e). For some 3D captures from the Bosphorus database, the extracted main nasal region is not fully connected, which mainly results from the noisy points or varying minor structure for some identities. Therefore, a morphological *closing* is applied to the thresholded SN_x to correct this.



(a) rotated capture (b) after alignment



(c) TSN_x (d) main nasal region (e) detected nasal bridge

Figure 7.3: Coarse roll rotation calibration using the detected nasal bridge: (a) An example of wrongly aligned captures from the Bosphorus database. (b) Results of calibration by the detected nasal bridge region. The candidate nasal bridge is first detected by TSN_x in (c) and the main nasal region shown in (d) is further extracted on the basis of the location of nose tip. The biggest connected area shown in (e) is then detected from (d). Using the estimated '*top*' (red rectangle) and '*bottom*' (green rectangle), the rotation angle (α_1) can be estimated, as describe in Section 7.2.1.2.

7.2.1.2 Coarse Roll Calibration

As can be seen from Figure 7.3(e), the detected nasal bridge is an irregular structure whose vertical deviation is hard to estimate. Linear regression provides an effective approach to detect the central line on such structure and can be directly applied to compute the rotation angle. However, this method can greatly increase the computation complexity, which is not desirable for the coarse roll rotation correction. Therefore, an alternative and efficient method is proposed in this section to rapidly estimate the rotation angle of the y axis (α_1) by finding the centroids, (top_x, top_y) and ($bottom_x$, $bottom_y$), of '*top*' (red) and '*bottom*' (green) of the nasal bridge.

The previously detected nose tip is used to help localize the '*top*' and '*bottom*' centroids. On the detected nasal bridge map shown in Figure 7.3(e), the position of the nose tip detected in the alignment phase can indicate the horizontal range and the '*bottom*' can be localized by the neighbouring points of the tip. The top of detected nasal bridge, shown as the red rectangle is used to estimate the '*top*'. Using the nasal bridge, all the pose variations in the roll direction have been approximately calibrated. As shown in Figure 7.3(b), the rotated capture $F_1(x_1, y_1, z_1)$ is computed from the original capture $F(x, y, z)$ with the rotation angle α_1 in the x-y plane.

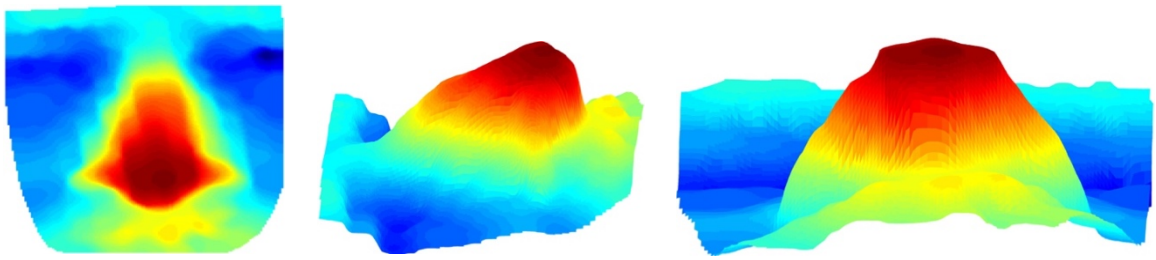
$$\alpha_1 = \tan^{-1} \frac{top_x - bottom_x}{top_y - bottom_y} \tag{4}$$

The new coordinates are then given by

$$\begin{cases} x_1 = x * \cos\alpha_1 - y * \sin\alpha_1 \\ y_1 = x * \sin\alpha_1 + y * \cos\alpha_1 \\ z_1 = z \end{cases} \tag{5}$$

7.2.1.3 Nose Tip Detection

The curvatures and SI are widely used for the nose tip detection as they can effectively detect the convex regions, with the tip located in the largest convex region. However, the main challenge of this method is how to accurately locate the tip within this region. In many landmarking algorithms, finding the centroid or the highest point is the most popular approach, which produces high accuracy and consistency tip's location. In fact, this method is not always reliable when some unexpected problems occur during the data acquisition or the target captures contain special nasal structures, which cannot be addressed by general denoising algorithms. Figure 7.4 gives an example of such kind of captures. Simply considering the depth information of nasal region might not necessarily produce an accurate tip position.



(a) Frontal view (b) Side view (c) Upward view

Figure 7.4: An example facial capture from the Bosphorus database (shown in three views) that suffers unexpected problems during the data acquisition, and fails in the tip detection when only the depth information is used.

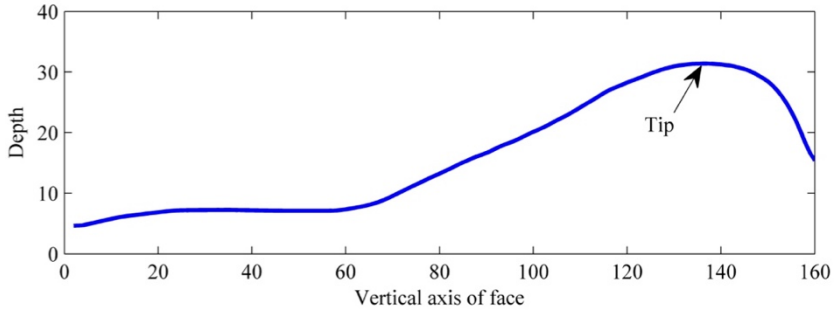


Figure 7.5: Using the nasal profile to detect the position of the nose tip in the y axis

It is hard to apply a generalized denoising approach to address such problems as they may cause other unaffected captures to lose fine details. Also, the detection of such captures is a challenging issue. Instead of further smoothing those captures, using the TSN_x map, for example that in Figure 7.3(c), and the nasal profile shown in Figure 7.5 can provide a novel and effective nose tip (Tip_x, Tip_y) detection approach. First, for a frontal view facial capture, the position of the nose tip in the x axis, Tip_x, is determined by the nasal bridge, which can be estimated from its neighbouring points on the nasal bridge by finding the centroid of the green box shown in Figure 7.3(e). Using this estimated Tip_x, the nasal profile is found by intersecting a plane that is perpendicular to the x-y plane, which results in a curve on the nasal surface, such as that shown in Figure 7.5. This curve is further smoothed to remove the noise and the position of the nose tip in the y axis, Tip_y, is located at the top of the profile.

7.2.2 Alar Grooves Detection and Self-Yaw Rotation Calibration

In most landmarking methods employing the depth map, the location of the alar grooves might suffer the influence caused by the noisy points or inaccurate nose tip position. In this section, the use of thresholded SNs aims to predict an accurate candidate alar grooves map and further correct the inaccurate tip and yaw rotations at the same time. Using the candidate adjusted maps, the alar grooves are localized on the depth map. The main steps are shown in Figure 7.6.

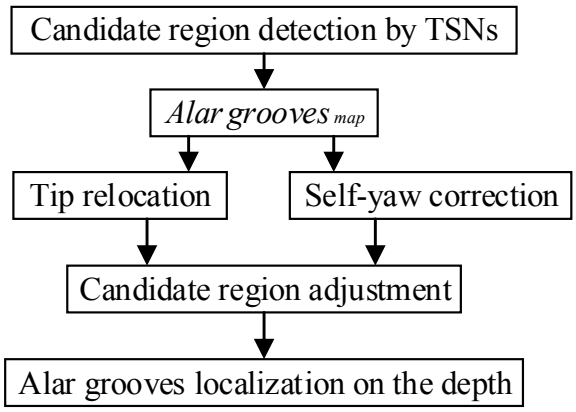
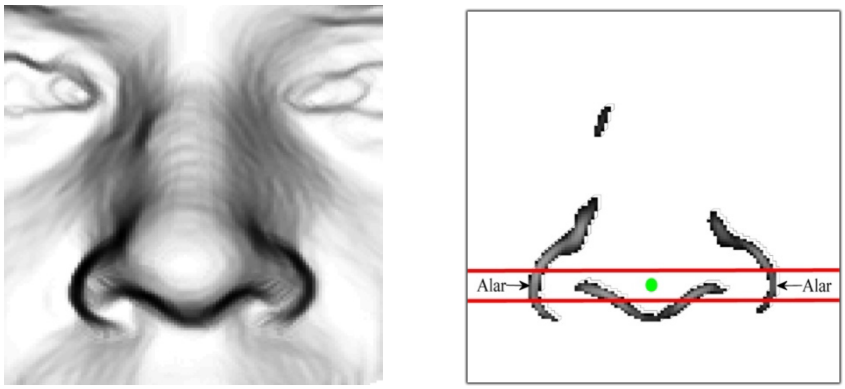


Figure 7.6: The flow chart of the proposed alar grooves detection approach.

7.2.2.1 Candidate Region Detection

Compared to other nasal and adjoining cheek regions, the SNs of nasal alar grooves and nostrils possess higher values in the SN_x and SN_y maps, which results in very small value in SN_z. Therefore, in the example SN_z map shown in Figure 7.7(a), the outline of the nasal alar grooves and nostrils is very salient in comparison with other regions around the nasal region and the candidate alar grooves can be obtained directly from the SN_z map. By thresholding the SN_z map, the initial candidate region is found, see Figure 7.7(b). This approach is more straightforward than using both the SN_x and SN_z maps [145]. Although some small patches on the upper nasal region are also included in the thresholded map, these can be further excluded by using the nose tip as a reference.



(a) (b)

Figure 7.7: Candidate alar grooves detection: (a) shows the SN_z map of nasal region and darker region denotes lower value. The alar grooves and nostrils are more salient than the other parts on the SN_z map. (b) The thresholded SN_z map used to find the candidate alar groove region. Based on the location of nose tip (green point), two horizontal red lines are used to determine the upper and lower boundary of the target region.

The $Alar\ grooves\ _{map}$, see Figure 7.7(b), is a matrix with the same size as SN_z and is given by

$$Alar\ grooves\ _{map} = \{Ag_{i,j} | Ag_{i,j} = f_z(i,j)\} \tag{6}$$

where, given SN_z,

$$f_z(i,j) = \begin{cases} 1 & if\ SN_z(i,j) < T_z \\ 0 & otherwise \end{cases} \tag{7}$$

T_z is set to 0.3 for candidate alar grooves thresholding. Using the location of the nose tip [green point in Figure 7.7(b)] detected in the previous part [7], two red lines are defined at a fixed vertical separation from the original nose tip, which is detected from the depth map.

The resulting horizontal region between the two red lines of height i pixels is regarded as the target region in which to find the candidate nasal alar grooves. In experiments, a suitable value has been found to be $i = 17$. For each line in this region, the initial candidate alar grooves points, $init\_Left_i$ and $init\_Right_i$, are obtained by finding the left and right extrema, respectively. However, self-occlusion caused by pose variations and inaccurate nose tip location greatly reduce the landmarking accuracy. To address this problem, nose tip relocation, self-yaw calibration and missing points interpolation are proposed.

7.2.2.2 Nose Tip Relocation

In the last section, the position of nose tip (J_{tip}) has been localized by the nasal bridge, which provides a coarse and reliable estimation. To further refine the previous detected nose tip and calibrate the remaining yaw rotations, the candidate alar grooves is proposed to generate a new nose tip location (J_new_{tip}). Combined with these two types of estimations of the nose tip, the rotation angle in yaw direction is estimated by their discrepancy.

The initial candidate points, $init\_Left_i$ and $init\_Right_i$, can be presented as i pairs of candidate points. Most pairs on the frontal view captures are relatively symmetrically distributed on each side of the nose. Therefore, it is possible to further refine the location of the nose tip using the detected pairs. Those pairs can be found by

$$\begin{cases} Dist\_left_i = J_{tip} - init\_Left_i \\ Dist\_right_i = init\_Right_i - J_{tip} \\ D_i = |Dist\_left_i - Dist\_right_i| \\ D_s = \min\left(D_i\ (Dist\_left_i \neq 0\ \cap Dist\_right_i \neq 0)\right) \end{cases} \tag{8}$$

where the distances between the initial candidate points ($init\_Left_i$ and $init\_Right_i$) and current nose tip (J_{tip}), $Dist\_left_i$ and $Dist\_right_i$, are first calculated. Then, the absolute distance, D_i, is obtained by calculating the inconsistencies between $Dist\_left_i$ and $Dist\_right_i$, which helps find the most appropriate and symmetric pairs, D_s, for the nose tip relocation.

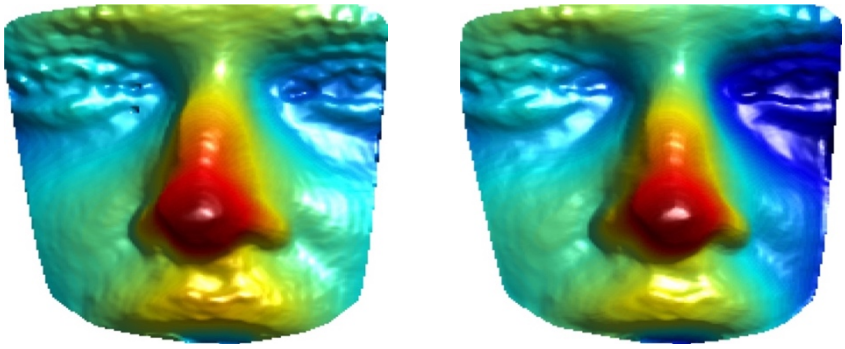
Some initial candidate pairs with an invalid distance are first removed and the pairs with minimum distance difference are used to predict the new nose tip. Also, there is no further adjustment of the nose tip when the value of D_s is smaller than T_s, which means the current nose tip only possesses a small offset when compared to the estimation of the points pairs. It is not necessary to address such deviations as the human nose is not strictly symmetric. However, if the value of D_s is bigger than T_s, which means the offset of current nose tip is very big, the current tip should be relocated by calculating the average distance of the existing set of pairs with low D_i, using

$$J\_new_{tip} = \begin{cases} \left\lceil \frac{1}{2n} * \sum_{i=1}^{n} init\_Right_s(i) + init\_Left_s(i) \right\rceil & D_S > T_s \\ J_{tip} & D_s \leq T_s \end{cases} \quad (9)$$

Therefore, using the candidate alar grooves points, nose tip locations with big deviations can be approximately calibrated and further used for yaw rotation calibration. Noisy data remaining on the facial surface might be the main reason for these less accurate tip locations. However, some other captures fail in the tip detection due to their intrinsic nose characteristics, for example of self-yaw rotated nose, which will be further explained and addressed in the next section

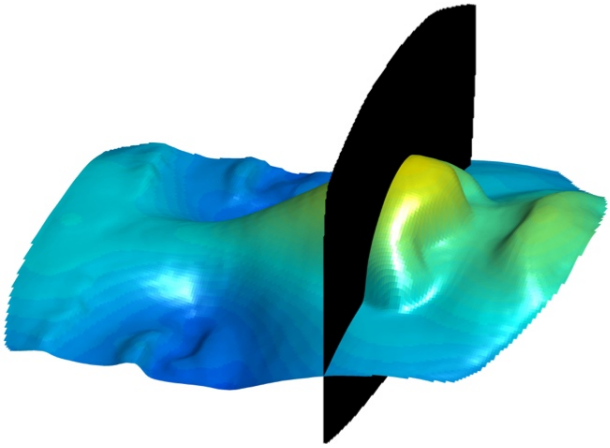
7.2.2.3 Self-yaw Rotation Calibration

For the majority of frontal view facial captures, the position of the nasal bridge is generally in the centre of the whole nasal region, which should also contain the nose tip. Therefore, if the estimated J_new_{tip} is not within the nasal bridge region, the captures must contain the yaw rotation. Actually, in the PCA based alignment, nearly all the pitch and yaw rotations have been successfully addressed. However, for around 10% of captures, some pose variations in yaw on the nasal region still remain. In addition to the basic facial pose variations, some yaw rotations result from special nasal anatomical structure variations for some identities, as shown in Figure 7.8(a). The main drawback of such structure is that the *x* axis might be reversed after PCA based alignment, which greatly decreases the within-class consistency of landmarking.

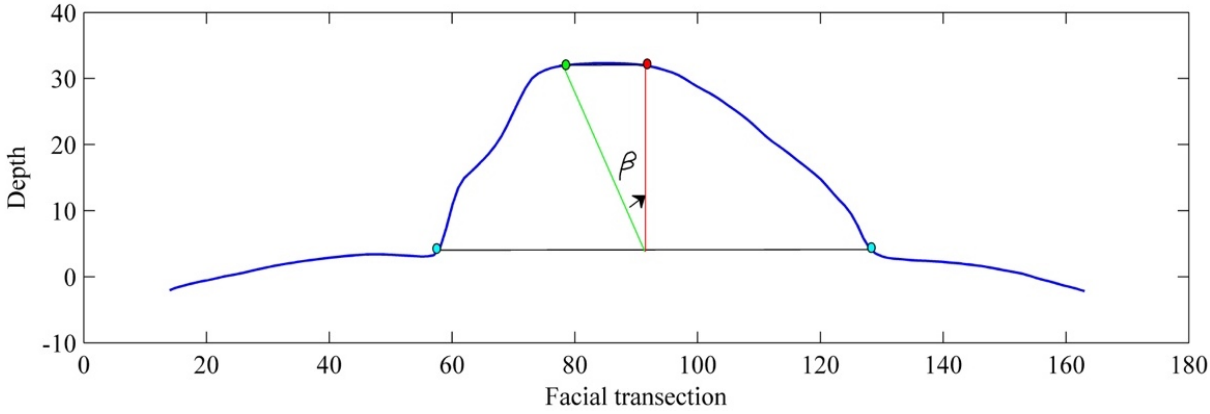


(a) Self-yaw rotation (b) Corrected nasal region

Figure 7.8: An example capture with self-yaw rotation from the Bosphorus database and its calibration results. The colour changes shown on the nasal region of (a) demonstrate that the nose suffers the yaw rotation, which is not caused by the facial pose variations as the cheeks regions are shown as frontal.



(a) Facial transection



(b) Self-yaw rotation estimation

Figure 7.9: An example of facial transection passing through the nose tip: Using planes that are perpendicular to the *x*-*y* plane results in a series of intersection curves shown in blue on the nasal surface. The rotation angle can be estimated by the detected points on this curve in terms of the tip (red) that is computed by the candidate alar grooves (lake blue) and bridge point (green).

This kind of yaw rotation presents more difficulties to the alar grooves localization: the rotated nose will cause the self-occlusion on one side of the nose corner and it is also hard to find the accurate features in the SNs. In general, most frontal view noses possess a relatively symmetric structure centred on the nose bridge, which can help design an effective landmarking strategy based on the symmetry. Therefore, such individuals with self-yaw variations might increase the computational complexity of the alar grooves detection.

To address the self-yaw rotation, the locations of the previous detected nose tip, J_{tip} and J_new_{tip}, are first applied to detect such rotations. Then, the detected nasal bridge shown Figure 7.3(c) and candidate alar grooves maps can help address the self-yaw variations. The nasal bridge is one of the most salient geometrical structures on the face surface, which is not be presented upright when face is rotated in the yaw direction. The detection of nasal bridge is complicated using depth information. However, the nasal bridge is more salient on the SN_x map and can be robustly detected by thresholding SN_x (TSN_x), even when the whole nasal region has self-yaw rotations.

When combined with the candidate alar grooves (*init_Right* and *init_Left*) matrix obtained in the previous section, which provide a reliable prediction of nose corners, the yaw rotation angle (β) is estimated by n valid pairs of corner candidate points and their corresponding nasal bridge points ($Nasal_{bridge}$) shown as green in Figure 7.9(b). The rotated capture $F_2(x_2, y_2, z_2)$ is computed from $F_1(x_1, y_1, z_1)$ with a rotation angle β in the x-y plane.

To calculate the rotation angle, n facial transections around the nasal tip are found by intersecting planes that are perpendicular to the x-y plane. Figure 7.9 illustrates an example of transection passing through the nose tip, resulting in an intersection on the 3D facial surface shown in blue. On this transection, there are two nose corners (lake blue) which are the candidate alar grooves, a bridge point (green) and the centre point (red). β is obtained by calculating the rotation angle from the bridge point to the tip by

$$\beta = \frac{1}{n} * \sum_{i=1}^{n} \tan^{-1} \frac{Nasal_{bridge}(i) - \frac{1}{2} * (init\_Right(i) + init\_Left(i))}{Depth\,(Nasal_{bridge}(i))} \tag{10}$$

and the rotated coordinates (x_2, y_2, z_2) are found by

$$\begin{cases} x_2 = x_1 * \cos\beta - z_1 * \sin\beta \\ y_2 = y_1 \\ z_2 = x_1 * \sin\beta + z_1 * \cos\beta \end{cases} \tag{11}$$

After the coarse roll and self-yaw calibrations, the nose tip is relocated using both the neighbouring points found on the nasal bridge and the prediction by the candidate alar grooves points.

7.2.2.4 Reconstruction of Missing Points and Candidate Region Adjustment

Noise or self-occlusion can result in some missing or false detected points in the candidate alar grooves region defined in Figure 7.7(b). To address these problems, using the newly adjusted nose tip (J_new_{tip}), the distances from the candidate points to the central point are recalculated and the missing points are reconstructed by the points from the same lines or interpolated by the neighbouring points, on the basis of the threshold T_c. According to the pairs previously used to predict the new nose tip, T_c is found by calculating the average distance from the candidate points to the new nose tip using,

$$\begin{cases} Dist\_left_i = J\_new_{tip} - init\_Left_i \\ Dist\_right_i = init\_Right_i - J\_new_{tip} \\ D_i = |Dist\_left_i - Dist\_right_i| \\ D_s = \min\left(D_i \left(Dist\_left_i \neq 0 \ \cap Dist\_right_i \neq 0\right)\right) \\ T_c = \left\lceil \frac{1}{m} * \sum_{i=1}^{m} Dist\_left_s(i) \right\rceil \end{cases} \tag{12}$$

where $Dist\_left_i$ and $Dist\_right_i$, are the distances between the initial candidate points and current nose tip. The absolute distance, D_i, is obtained by measuring the difference between $Dist\_left_i$ and $Dist\_right_i$, which helps find the most appropriate and symmetric pairs (D_s). s indicates the selected pairs for the missing points reconstruction.

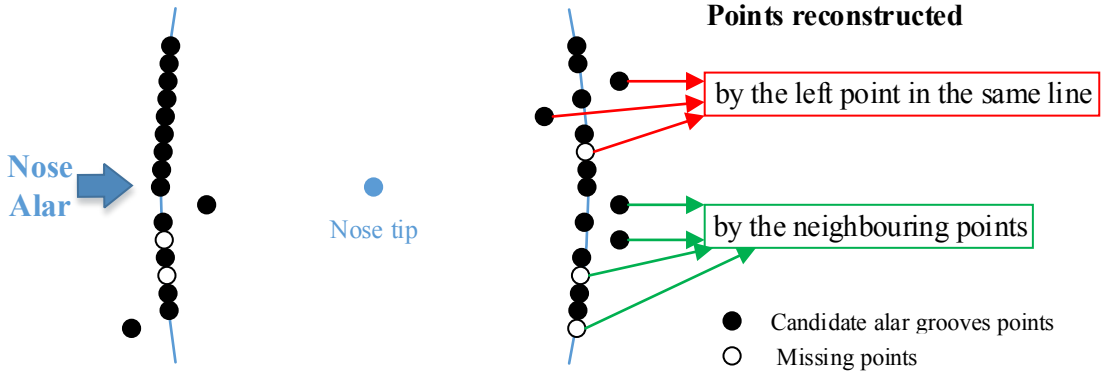


Figure 7.10: The diagram describes the missing points reconstruction and candidate alar grooves adjustment. For each missing or false detected point on the right alar, $Dist\_right_i$, they are reconstructed on the basis of the pair on the same line. If the pair on the left is well detected, which is measured by T_c, the point on the right can be reconstructed directly by this pair and the location of nose tip. Otherwise, the point on the right should be interpolated by the neighbouring points.

For each missing or false detected point, the distance between its symmetrical point in the same line and the nose tip is first compared with T_c, which determines whether it is reconstructed by the pair in the same line or neighbouring points. The correction methods of different kinds of missing or false detected points are described in Figure 7.10.

7.2.2.5 Alar Grooves Localization on the Depth Map

By using the thresholded SNs maps and the location of the nasal bridge, the resulting alar grooves provide an accurate reference map for final alar grooves landmarks on the depth map. Finally, the depth information is further applied to produce more accurate alar grooves on the 3D facial surface. More specifically, the candidate alar grooves points shown in lake blue are first localized on the facial transection, as shown in Figure 7.9(b), which produce an approximate location on the surface normal maps.

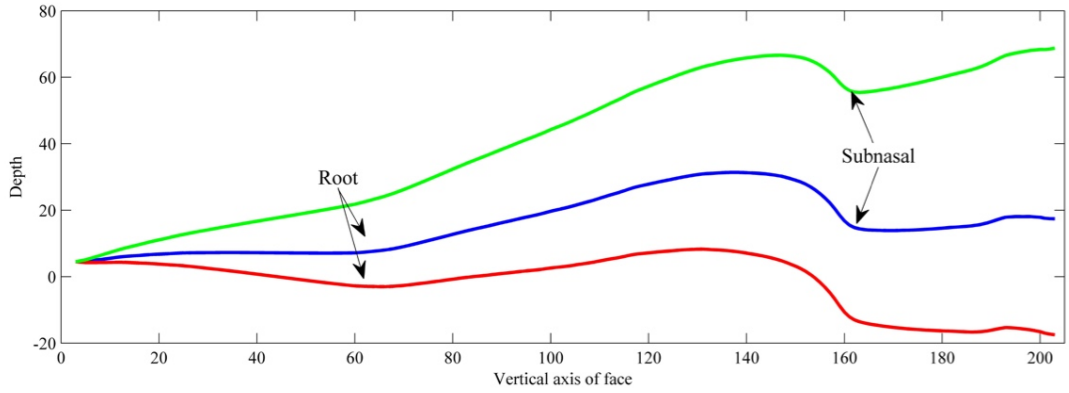
To refine the alar grooves more locally on the depth map, the facial surface curve resulting from intersecting the plane, which is perpendicular to the *x*-*y* plane, is used. On the facial curves shown in Figure 7.9(b), the two alar grooves should be located at the inflexion points on the facial curve. In a similar manner to the nose tip localization shown in Figure 7.5, the most effective way to detect these points is using the rotated curves. Specifically, from the nose tip, the facial curve is first divided into two halves and the left and right half are then rotated clockwise and counter-clockwise in the *x*-*y* plane, respectively. The two alar grooves are then located at the local minima of two halves.

7.2.3 Fine Roll Rotation Calibration and Root Localization

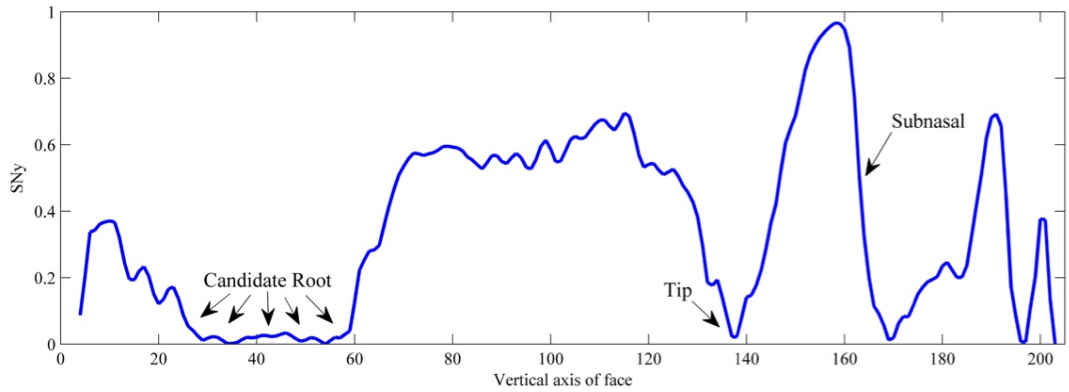
7.2.3.1 Nasal Root and Subnasal Localization

Nasal root localization is the most challenging step in the landmarking process. It is hard to define an accurate root location as the root area is relatively flat for some types of human nose. Using a reference area provided by thresholded surface normal maps can successfully estimate the candidate nasal root position. However, simply calculating the centre of this area might not produce an accurate root position [145]. Therefore, using a similar approach to the nose tip adjustment, the position of nasal root and subnasal in *x* axis are first narrowed by using the detected nasal bridge shown in Figure 7.3(d), as these two landmarks should be located within the nasal bridge. The fusion of both thresholded SNs maps and nasal bridge produces a relatively accurate horizontal positions of the root and subnasal, respectively.

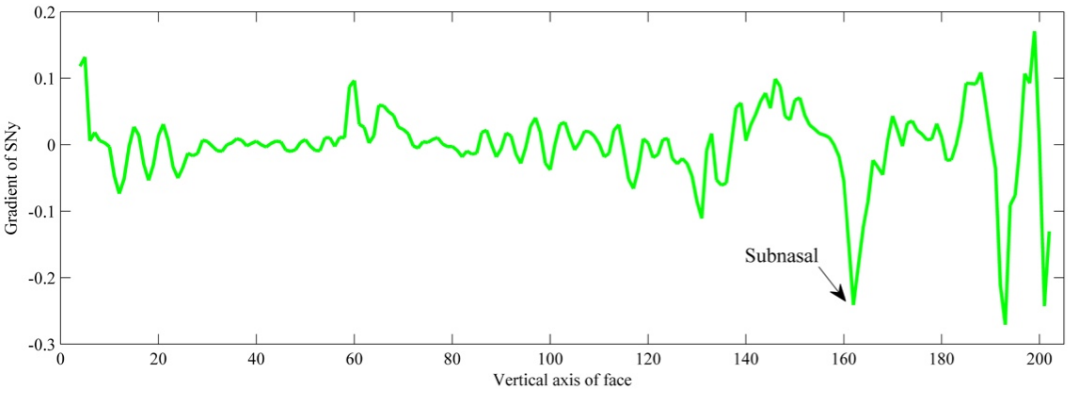
For their positions in y axis, the facial profile curves which are generated on the basis of the horizontal prediction by the nasal bridge are proposed to draw on the depth and SN_y maps in Figure 7.11.



(a) Nasal profile signature on the depth map: the blue curve is the original profile signature, whose clockwise (red) and counter clockwise (green) rotation are generated to localize the root and subnasal, respectively.



(b) Nasal profile signature on the SN_y map. The nasal root is located on the top of nasal bridge, which is relatively flat. The nose tip presents as the inflexion point on this curve. The subnasal locates at the fast decreasing point.



(c) Gradient of the blue curve in (b)

Figure 7.11: Root and subnasal localization using facial profile signatures on the Bosphorus database.

These two landmarks should locate at the local minima of facial profile curves on the depth map. However, it is hard to find the local minima on the profile curves of certain captures, for example of the blue curve shown in Figure 7.11(a). To address this problem, the clockwise rotated curve with the rotation angle 10˚ shown in red in Figure 7.11(a) has been proposed to find the local minima and finally used to localize the root [143]. In a similar manner, the subnasal is located at the local minima of the counter clockwise rotated green curve with rotation angle 15˚, as shown in Figure 7.11(a).

According to the structure of nasal region, the nasal root is located on the top of nasal bridge, which is relatively flat on the SN_y map and the subnasal locates at the fast decreasing point. Therefore, combined with the position provided by the SN_y map as shown in Figure 7.11(b), the nasal root and subnasal can be localised at the inflexion points in Figure 7.11(a). The only difference is that the reference area of subnasal cannot be directly provided by the SN_y maps, which is presented as the fast decreasing point on the profile signature curve obtained from SN_y map, as shown in Figure 7.11(b). Therefore, the gradient of the SN_y profile signature curve is generated to help localize the subnasal, as illustrated in Figure 7.11(c), where the subnasal is clearly marked as local minimum.

Therefore, all of those significant steps mentioned above in terms of thresholded surface normal maps, the detected nasal bridge and nasal profile signatures produce a relatively accurate prediction in both vertical and horizontal position of the nasal root and subnasal.

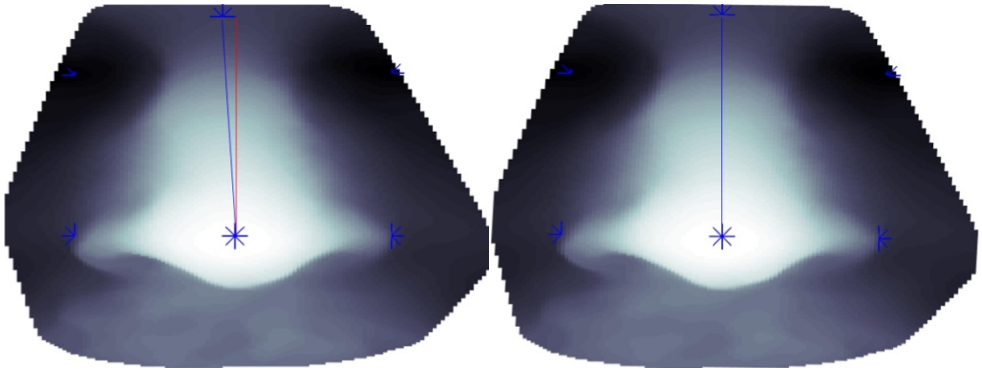
7.2.3.2 Fine Roll Rotation Adjustment by Landmarks

As the detected nasal bridge region is an irregular structure, it can only help provide an approximate rotation angle and a rough pose alignment. In some captures, such as the one shown in Figure 7.12(a), there is slight roll rotation that can be effectively detected by the locations of the nasal root and tip. A simple way to address this problem is using the detected root and tip, shown as blue points in Figure 7.12(a), to calculate the rotation angle (α_2) between the blue line from root to tip and the y axis (shown in red). As shown in Figure 7.12(b) the rotated capture $F_3(x_3, y_3, z_3)$ is computed with α_2 in the x-y plane from $F_2(x_2, y_2, z_2)$ by

$$\alpha_2 = \tan^{-1}\frac{root_x - tip_x}{root_y - tip_y} \tag{13}$$

and the rotated coordinates (x_3, y_3, z_3) are found by

$$\begin{cases} x_3 = x_2 * \cos\alpha_2 - y_2 * \sin\alpha_2 \\ y_3 = x_2 * \sin\alpha_2 + y_2 * \cos\alpha_2 \\ z_3 = z_2 \end{cases} \tag{14}$$



(a) facial capture with slight roll rotation (b) aligned capture

Figure 7.12: Roll rotation calibration using the detected landmarks.

7.2.4 Eye Corners Detection

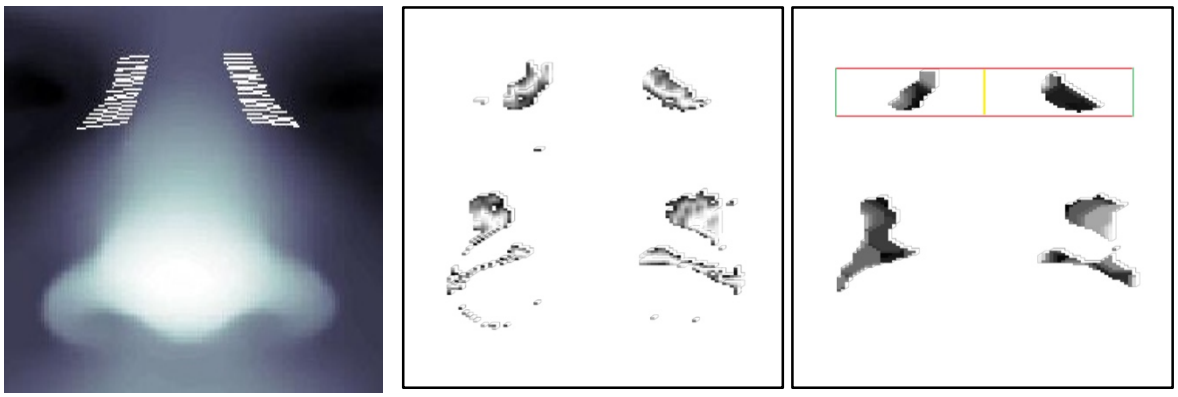
7.2.4.1 Eye Regions Localization by Thresholded SNs

In the first stage, the approximate eye region in horizontal is found by thresholding surface normal maps. The concave region on both sides of nasal root, as the white lines shown in Figure 7.13(a), possess distinct geometry structure from other parts around the nasal root on the depth and surface normal maps. Instead of calculating and thresholding the curvatures, simply thresholding SNs can also produce a reliable candidate eye corner (*CEC*) position. The surface normal on this region is partial to *x* and deviates from *y* direction, which brings the possibility to find the regions shown in Figure 7.13(b), according to the following equation. T_x and T_y are set to 0.7 and 0.15.

$$CEC = \{C_{i,j} | C_{i,j} = f_{x,y}(i,j)\} \tag{15}$$

Where given SN_x and SN_y

$$f_{x,y}(i,j) = \begin{cases} 1 & (if\ SN_x(i,j) > T_x) \wedge (SN_y(i,j) < T_y) \\ 0 & otherwise \end{cases} \tag{16}$$



(a) Target concave region around the root (b) Thresholded SNs (c) SNs after postprocessing

Figure 7.13: Candidate eye corner localized by the thresholded SNs

As shown in Figure 7.13(b), because of the noisy points, the detected *CEC* region might not be connected in some parts. Therefore, the morphological *close* is first applied on the initial *CEC* using a disk-shaped structuring element with radius 5. Also, the morphological *erosion* is employed to remove some small noisy points. Figure 7.13(c) illustrates the final candidate region after post-processing.

7.2.4.2 Corner Localization

Using the detected *CEC* region as well as the alar grooves found above, the final eye corner region can be fixed in the rectangular region. As shown in Figure 7.13(c), the top and bottom boundary (red lines) are found according to the *CEC* region and the left and right ones (green lines) are determined by the detected alar grooves. The whole final eye corner region can be divided into two sides on the basis of the nasal root and the eye corners are located in the global minimum of each side.

7.3 Landmarking Performance Evaluations

7.3.1 Within-class Consistency Evaluation

Although the Bosphorus database provides some manually detected landmarks, finding the ground truth locations for these landmarks is still very challenging. Evaluating the within-class consistency is a widely used strategy, which is verified by calculating the Euclidean distance between each landmark for captures from the same subject. To be more specific, all the captures are pose aligned and translated to the nose tip and the mean and standard deviation (std) of the distance (in mm) of the detected landmarks for all the subjects are illustrated in Table 7.1. These results demonstrate that the proposed landmarking method possess high within-class consistency in comparison with [7].

In a similar manner, the landmarks' consistency is also estimated using the FRGC v.2 database as this version contains more expression variations in comparison with Spring 2003 folder. As can be seen from Table 7.1, the within-class consistency of nasal root has been significantly improved. For comparison with [7], the subnasal evaluation is not included in this evaluation.

Using the landmarks' within-class consistency can determine whether the target landmark is localized in the same point or area for each capture of the same identity. It does produce a good reference for landmarking evaluations, especially for the neutral captures or captures with very few variations on the nasal region. However, in real world application scenarios, most captures have expressions which might cause facial surface movements. Therefore, the target landmarks

might shift as the shape of the nasal region changes, and it is not appropriate to evaluate the landmarking by the widely used within-class consistency calculation method in these cases.

To further explain the drawback of current within-class consistency evaluation methods and demonstrate the high consistency of the proposed landmarking algorithm when the captures contain less nasal region variations, the within-class consistency is further computed using different types of expressions from the Bosphorus database. The Bosphorus database contains various types of expressions, including 6 basic expressions, the expressions occur on different face action units (UFAU or LFAU) and their combinations (CAU). On the basis of the labels provided in the database, all the captures from 105 subjects are divided into 34 groups as shown in Table 7.2 and the consistency of the nasal root and two alar grooves are calculated in each group. The neutral capture of each identity is applied to compute the consistency.

As can be seen from Table 7.2, different expression variations on the nasal region have more or less influence on the landmarking consistency. For example, upper face action units (UFAU), which cause the nasal root area to stretch or extrude, show poor consistency of nasal root. In contrast, the lower face action units which exhibit movement in lower parts of face demonstrate poor consistency for the alar grooves, especially for the captures contain happy expression. All of those confirm the drawback of using the consistency to evaluate the landmarking performance of some non-neutral captures.

Table 7.1: A comparison of within-class landmarks consistency tested on the Bosphorus and FRGC v.2 databases. All the captures are pose aligned and translated to the nose tip and the mean and std of the distance of the detected landmarks for all the subjects are illustrated.

Database	Landmarks	Emambakhshm *et al*. [7]		Proposed	
		Mean	Std	Mean	Std
Bosphorus	Root	3.72	2.40	1.79	1.53
	Alar_L	2.06	1.50	1.31	1.03
	Alar_R	2.06	1.53	1.32	1.03
FRGC	Root	3.27	2.10	1.84	1.15
	Alar_L	1.29	0.98	1.24	0.89
	Alar_R	1.29	0.92	1.17	0.89

Table 7.2: Within-class landmarks consistency evaluation using different types of expressions. All the captures contain the expression variations from the Bosphorus database are divided into 34 groups and the mean distance of the landmarks for all the subjects in each group is calculated.

Types of expressions	Mean			Types of expressions	Mean		
	Root	Alar_L	Alar_R		Root	Alar_L	Alar_R
CAU_A12A15_0	1.81	1.75	1.45	LFAU_17_0	1.42	1.60	1.39
CAU_A22A25_0	1.33	1.30	1.05	LFAU_18_0	1.57	1.40	1.28
CAU_A26A12lw_0	1.29	1.18	1.15	LFAU_20_0	1.63	1.40	1.40
E_ANGER_0	2.02	1.27	1.13	LFAU_22_0	1.66	1.33	1.24
E_DISGUST_0	2.68	1.57	1.69	LFAU_23_0	1.64	1.33	1.27
E_FEAR_0	1.55	1.33	1.14	LFAU_24_0	1.50	1.32	1.26
E_HAPPY_0	1.54	2.67	2.29	LFAU_25_0	1.18	1.10	1.07
E_SADNESS_0	1.79	1.63	1.30	LFAU_26_0	1.29	1.31	1.13
E_SURPRISE_0	1.40	1.25	1.27	LFAU_27_0	1.69	1.28	1.35
LFAU_10_0	2.26	1.95	1.84	LFAU_28_0	2.15	1.57	1.45
LFAU_12LW_0	1.28	1.07	1.39	LFAU_34_0	1.76	1.28	1.27
LFAU_12L_0	1.34	1.41	1.57	LFAU_9_0	4.52	1.67	2.03
LFAU_12R_0	1.47	1.44	1.34	UFAU_1_0	2.05	1.18	1.31
LFAU_12_0	1.50	1.95	2.26	UFAU_2_0	1.61	1.15	1.04
LFAU_14_0	1.44	1.53	1.32	UFAU_43_0	1.52	1.12	1.14
LFAU_15_0	1.46	1.63	1.27	UFAU_44_0	1.97	1.16	0.98
LFAU_16_0	1.30	1.11	0.99	UFAU_4_0	2.25	1.31	1.28

Compared to captures containing nasal region variations, the proposed landmarking algorithm demonstrates good within-class consistency for the landmarks that have less surface changes. These include most of the alar grooves detected in UFAU groups, the root detected in LFAU groups and some alar grooves detected in LFAU groups. Some types of expression which cause very few facial movements, for example LFAU_16_0 and LFAU_25_0, provide a promising within-class consistency.

Therefore, the widely used within-class consistency evaluation method is not an appropriate manner to demonstrate the landmarking accuracy, when the subjects contain both the neutral and non-neutral captures. For the captures contain expression variations, the successfully detected landmarks might not produce a good consistency due to the facial surface movement. The measurement of such movement by a generalized method is a challenging issue. The

recognition performance evaluation on the features extracted by the detected landmarks provides a reasonable and effective way to indicate the landmarking accuracy, which will be further discussed in the next section.

Emambakhshm *et al*. [143] proposed a new within-class consistency calculation method, which is obtained by calculating the standard deviation. For each given landmark, the standard deviations of *x*, *y* and *z* axis are first estimated and their mean standard deviation is considered as the consistency. Using this calculation method, a comparison of the within-class consistency between the proposed landmarking and the landmarking explained in [143] is shown in Table 7.3.

The results demonstrate that the proposed landmarking algorithm is more consistent than that illustrated in [143]. The within-class consistency of all the detected landmarks have been significantly improved, especially for the captures in the FRGC database.

Table 7.3: A comparison of within-class landmarks consistency tested on the Bosphorus and FRGC databases using the method proposed in [143]. All the captures are pose aligned and translated to the nose tip and the mean and std of the standard deviations of the detected landmarks for all the subjects are illustrated.

Database	Landmarks	Emambakhshm *et al*. [143]		Proposed	
		Mean	Std	Mean	Std
Bosphorus	Root	1.06	0.58	0.75	0.20
	Alar_L	1.06	0.62	0.49	0.12
	Alar_R	1.19	0.60	0.48	0.12
	Eye_corner_L	1.76	1.03	1.03	0.22
	Eye_corner_R	2.12	1.14	1.01	0.22
	Subnasal	1.11	0.38	0.58	0.21
FRGC	Root	2.04	1.09	0.55	0.32
	Alar_L	1.29	0.82	0.29	0.20
	Alar_R	1.22	0.62	0.31	0.21
	Eye_corner_L	2.95	1.61	0.87	0.48
	Eye_corner_R	2.91	1.53	0.91	0.51
	Subnasal	1.86	0.85	0.47	0.29

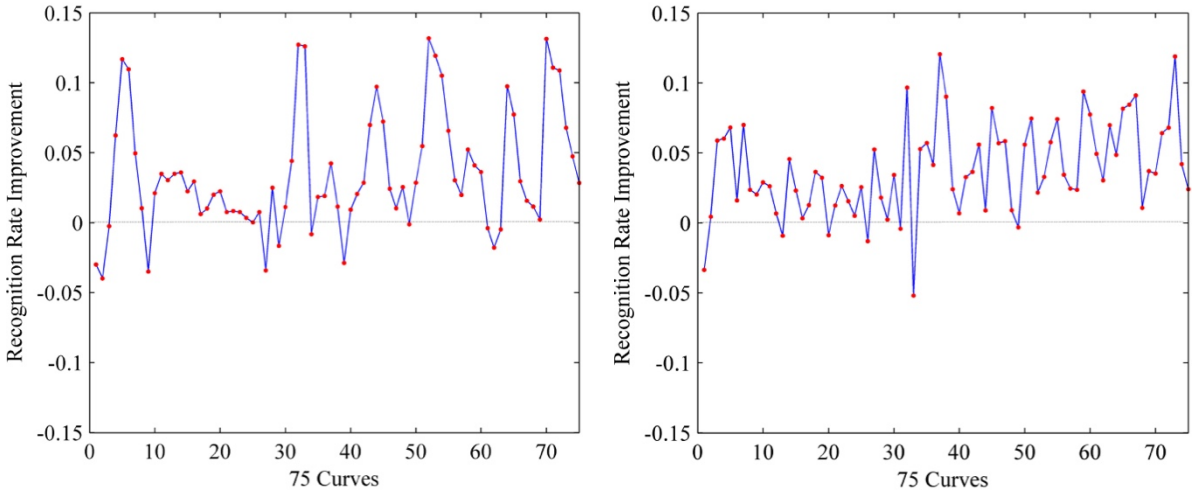
7.3.2 Recognition Performance Evaluation

In addition to using the within-class consistency, the effectiveness of these newly detected landmarks can also be evaluated by using the recognition performance, especially for the captures containing nasal region variations. Using the location of nasal root, tip and alar grooves, the Nasal Curves Matching (NCM) algorithm proposed in [7] is applied to extract 75 nasal curves to build the feature space. In NCM, each curve was resampled to 50 points. Gao *et al*. [34] used only 15 points per curve instead of 50 points and proved that fewer points on the nasal and cheek region can also produce an acceptable recognition performance, even with the dimensionality of the feature space greatly reduced. The 75 nasal curves with 15 points per curve that were used to extract discriminative features are shown in Figure 6.1, and these are used to evaluate the recognition performance of the new landmarking algorithm.

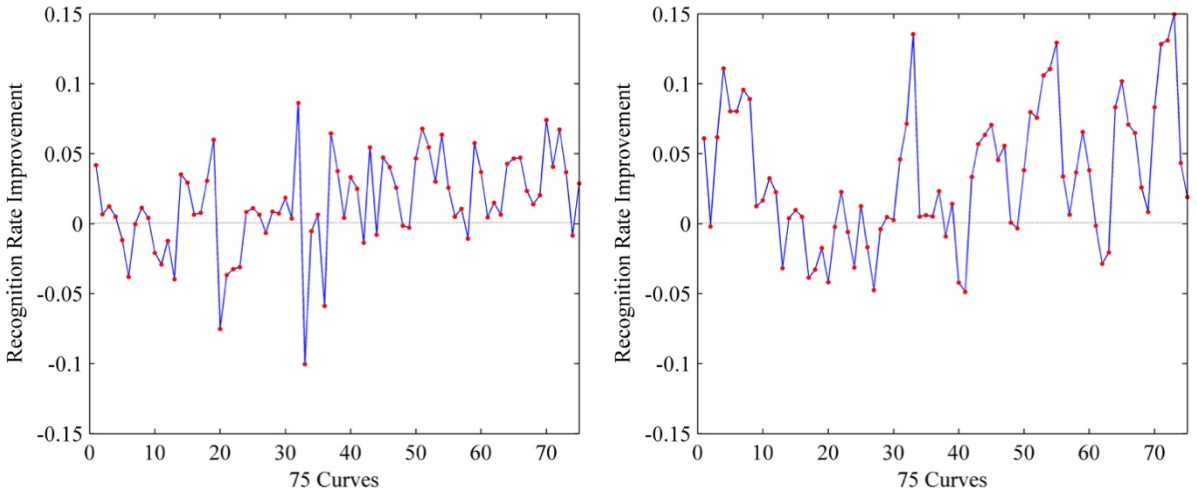
Using the leave-one-out (LOO) cross validation with city-block distance, the recognition rate of each single curve is obtained. To demonstrate the recognition performance improvement of the proposed landmarking algorithm, the method proposed in [7] is used for comparison. In Figure 7.14, the recognition rate improvement of each curve is calculated by the subtraction of recognition rate obtained from that of the proposed algorithm [7]. Positive value means the recognition performance using the proposed one is higher. The four components, Depth, SN_x, SN_y and SN_z, are evaluated individually.

As can be seen from Figure 7.14, the recognition performance of most curves extracted from Depth (63 curves) and SN_x (68 curves), and some curves on the SN_y (54 curves) and SN_z (55 curves) are improved by using the proposed landmarking algorithm. The recognition rate of some curves can increase by ~15%, in particular on the Depth and SN_z maps. In general, the recognition performance has been greatly improved using the proposed method and curves extracted from Depth and SN_x experience greater increase. There are very few curves suffered about 5% decrease, especially for{ features extracted from SN_y and SN_z maps.

To reduce the dimensionality and find the most discriminative set of curves, forward sequential feature selection is applied to the 75 curves, combined with the leave-one-out cross validation and nearest neighbour classifier. The R_1RRs obtained from different components of the 3D data as the size of the feature set increases, determined by the number of curves selected, are shown in Figure 7.15. Compared to the NCM [7], the recognition performance of all the components (Depth, SN_x, SN_y and SN_z) has been significantly improved by the proposed landmarking algorithm.



(a) Depth (b) SN_x



(c) SN_y (d) SN_z

Figure 7.14: Recognition performance improvement of each single curve using the proposed landmarking algorithm on the Bosphorus database. The red dots in each curve indicate recognition rate improvement in comparison with using the landmarks in NCM algorithm [7]. Positive recognition rate improvement means the recognition performance using the proposed landmarking is higher.

Given the best combination of curves selected by FSFS in Figure 7.15 for each component, the recognition performance is further evaluated under both identification and verification scenarios using these selected curves. The equal err rate (EER) is used for verification and the R_1RR for identification. Kernel Fisher's Analysis [139] with the polynomial kernel is used to calculate the similarity between the gallery and probe captures and the cosine distance is used in the feature matching.

In the Bosphorus database, nearly all the captures contain different degrees or types of expressions and thus evenly dividing them into the training and testing sets guarantees that the

various expressions are evenly distributed in both subsets. As this work aims to find the most discriminative and expression robust features, only those captures without large pose variations or occlusions are considered. The number of curves and the recognition rate they produce are illustrated in Table 7.4. The recognition performance of the four components tested under both scenarios has been significantly improved by the new landmarking method, especially for Depth and SN_x, where the improvement is 8-10%.

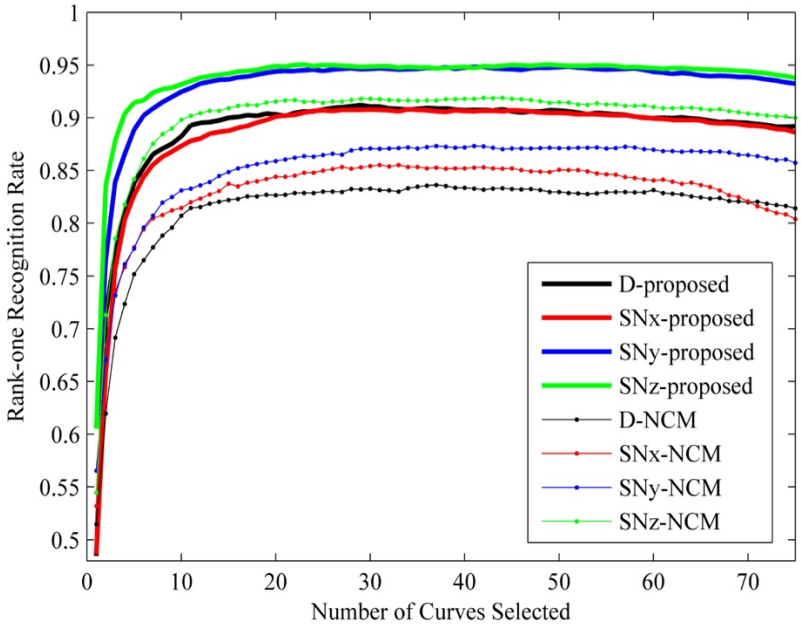


Figure 7.15: Rank one recognition rate against the number of nasal curves selected by the FSFS algorithm on the Bosphorus database, using both the proposed landmarking algorithm shown in bold and NCM [7] shown in the line with dots.

Table 7.4: Recognition performance evaluation using selected curves combination for four components under both identification and verification scenarios on the Bosphorus database. LOO cross validation and KFA classifier are used for feature selection and recognition performance evaluations, respectively.

		Depth	SN_x	SN_y	SN_z
NCM [7]	No. of Curves selected by LOO (R_1RR)	37 (83.64%)	31 (85.52%)	37 (87.33%)	42 (91.89%)
	R_1RR (KFA)	84.02%	82.24%	91.72%	93.79%
	EER (KFA)	5.78%	6.92%	3.07%	2.85%
Proposed	No. of Curves selected by LOO (R_1RR)	29 (91.19%)	34 (90.81%)	38 (94.79%)	23 (95.03%)
	R_1RR (KFA)	92.51%	92.29%	95.29%	95.22%
	EER (KFA)	3.21%	3.27%	1.71%	2.07%

The recognition performance has also been evaluated on the FRGC v.2 database, which produces similar results to the Bosphorus captures. The proposed landmarking algorithm clearly outperforms the NCM in terms of the single curve comparison, LOO cross validation for curves selection and recognition performance evaluation under identification and verification scenarios.

7.4 Landmarking Using Reconstructed 3D Captures from Photometric Stereo

As shown in Figure 7.1(a), the original outputs of the Photoface device is four images with different light directions. The surface normals and albedo with high resolution and accuracy shown in Figure 7.1(b)-(e) can be calculated from these four images on the basis of the Lambert's law. Compared to the depth maps reconstructed from 3D point clouds in other databases, the depth maps of Photoface captures as shown in Figure 7.1(f) is reconstructed from the SNs. The nasal landmarking approaches that successfully applied to the higher accuracy depth maps might not necessarily achieve good performance using relatively lower accuracy depth maps reconstructed from surface normals as many fine details and local minor structures on the nasal region for feature detection are missing, especially for employing the near infrared light.

Instead of using the smoothed depth maps, the resulting dense field of SNs and albedo from PS device might well suitable for the feature detection on the Photoface captures. Therefore, to avoid the loss of useful local features during the reconstruction and inspired by the landmarking approach using SNs in the previous sections, an alternative landmarking approach is proposed by utilising the SNs and albedo under near infrared light. The main advantage is that SNs can provide better representations of the 3D captures from photometric stereo, as more significant information is missing during depth reconstruction (see Figure 7.1). In addition, it is beneficial for the real world applications, as the reconstruction of depth information increases the computation complexity.

7.4.1 Nose Tip and Root Localization

7.4.1.1 Candidate Region Detection by Thresholding the Surface Normals

Nose tip detection is a vital and initial step in nasal landmarks localization, which provides the reference position for other nasal landmarks. However, for the Photoface captures, the

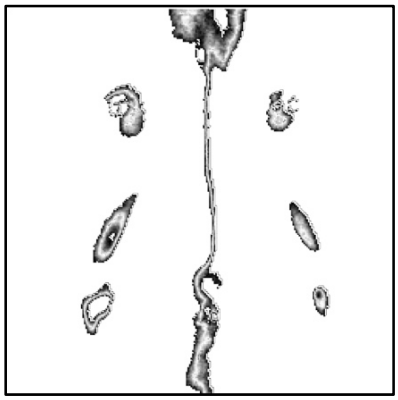
reconstructed depth map is not used in the landmarks localization, which means that the landmarking diagram proposed in Figure 7.2 cannot be applied directly as no initial tip position provided for other landmarks localization. Therefore, the nose tip should be first detected in the landmarking of the Photoface captures. Compared to other parts on the face, the nose tip is also easy to localize as its geometrical properties are salient in most 3D representations, including depth, surface normals, shape index and curvatures. For example on the SNs, the nose tip and its neighbouring points present higher values on the SN_z map and very small values on the SN_x and SN_y maps, which help localize the nose tip by thresholding three components of SNs.

In addition, the upper part of the nasal root has similar properties on the surface normals maps. Therefore, both the candidate nose tip (C_{Tip}) and root (C_{Root}) region can be found by

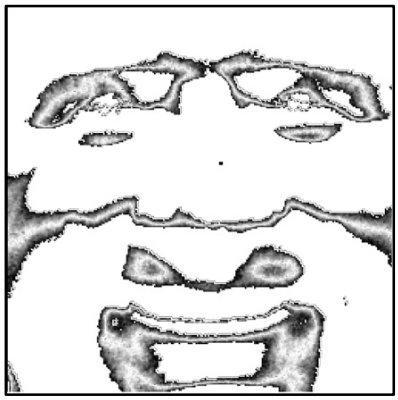
$$C_{Tip}\ and\ C_{root} = \left\{TR_{i,j} \middle| TR_{i,j} = f_{x,y,z}(i,j)\right\} \tag{17}$$

where, given SN_x, SN_y and SN_z,

$$f_{x,y,z}(i,j) = \begin{cases} 1 & if\ (SN_x(i,j) < T_x) \wedge (SN_y(i,j) < T_y) \wedge (SN_z(i,j) < T_z) \\ 0 & otherwise \end{cases} \tag{18}$$



(a) Thresholded SN_x (TSN_x)



(b) Thresholded SN_y (TSN_y)

(c) Thresholded (TSN_z) (d) $TSN_x \cap TSN_y \cap TSN_z$

Figure 7.16: The candidate nasal tip (red rectangle) and root (green rectangle) region localization on the Photoface captures by thresholding three components of surface normals. An example capture from the near inferred light part of the 3DE-VISIR database is used for demonstrating the resulting map of each component after thresholding.

The thresholds for the captures using near infrared light, T_x, T_y and T_z, are set to 0.1, 0.1 and 0.98, respectively. The surface normals thresholding results of an example capture from the 3DE-VISIR database by the near infrared light are illustrated in Figure 7.16(a)-(c) demonstrate the selected facial patches of each component and (d) is the result of the combining three components.

As shown in Figure 7.16(d), thresholding the surface normals can quickly eliminate most irrelevant regions on the face, except some convex regions on the month, cheek and eye areas that possess similar characteristics in surface normal maps. The candidate nasal root and tip areas are successfully detected shown as the green and red, respectively. However, to localize the tip root, all the candidate regions should be further refined as this is considered in the next section.

7.4.1.2 Nasal Tip and Root Localization by SNs and Gradient Calculation

The nasal bridge, as a salient structure centred on the human face, narrows the vertical location of the detected candidate region, which can quickly eliminate the irrelevant parts around the eye and cheek regions. Therefore, the nasal bridge is first detected before the nasal tip and root localization. Generally, the characteristic of the nasal bridge is distinct on the human face, especially in the SN_x. For the captures with higher accuracy depth maps, using the threshold SN_x is an effective approach, with the reference provided by the nose tip. However, it is hard to detect the bridge directly in TSN_x [Figure 7.16(a)] as there is no initial tip information provided.

Therefore, the calculation of the gradient of SN_x (gx) as shown in Figure 7.17(a), is proposed for the bridge detection, as the sign changes occur on the bilateral nasal bridge and also the gradient is bigger than the other nasal parts. The candidate bridge is found by the thresholded gx map as shown in Figure 7.17(b). Using the morphological processing, regions in Figure 7.17(b) are vertically grouped to several parts and the largest connected region is found by K-means clustering, which is the nasal bridge.

The horizontal location of the candidate region can be narrowed on the basis of the special properties of the subnasal and mouth areas. The subnasal and mouth areas have higher SN_y values than the root and nasal bridge, as they contain some convex and concave regions with the peaks and sign changes. Therefore, using the thresholded SN_y map shown in Figure 7.17(c), the boundary of the subnasal and mouth area is found, which can help eliminate the irrelevant regions in Figure 7.16(d) and provide an appropriate prediction of the candidate tip region.

(a) gradient of SN_x (gx) (b)thresholded gx (c) thresholded SN_y

Figure 7.17: Nasal bridge detection on the Photoface captures by thresholding the gradient of SN_x and SN_y maps. An example capture from the near infrared light part of the 3DE-VISIR database is employed.

Combined with both the horizontal and vertical prediction, it is easy to differentiate the candidate nasal root (green) and tip (red) region as shown in Figure 7.16(d). The nose tip is located in the centre of box. For the nasal root, the candidate region in the green box provide a relatively accurate position of the root in horizontal, whose lower boundary denotes the root. Although the approximate region has been found in the green box, it is still challenging to localize it vertically as some deviations might cause big changes in SN_x map, even changing the signs. To address this instability, the vertical position of the nasal root is estimated by the neighbouring points on the nasal bridge.

7.4.1.3 Nasal Tip and Root Localization by Albedo Central Profile

In addition to the use of thresholded SN_y map shown in Figure 7.17(c), the vertical position of the nasal root and tip can be found by the central profile of the human nose on the albedo map as shown in Figure 7.18, which provides the approximate vertical distribution of the hair, nasal root, nasal tip and mouth. The central profile is found on the basis of the estimated centre of the nasal bridge detected in Figure 7.17(b). Combined with the candidate region found in

Figure 7.16(d), some extrema points shown in Figure 7.18, which indicate the root, tip and month, are found to predict the final positions.

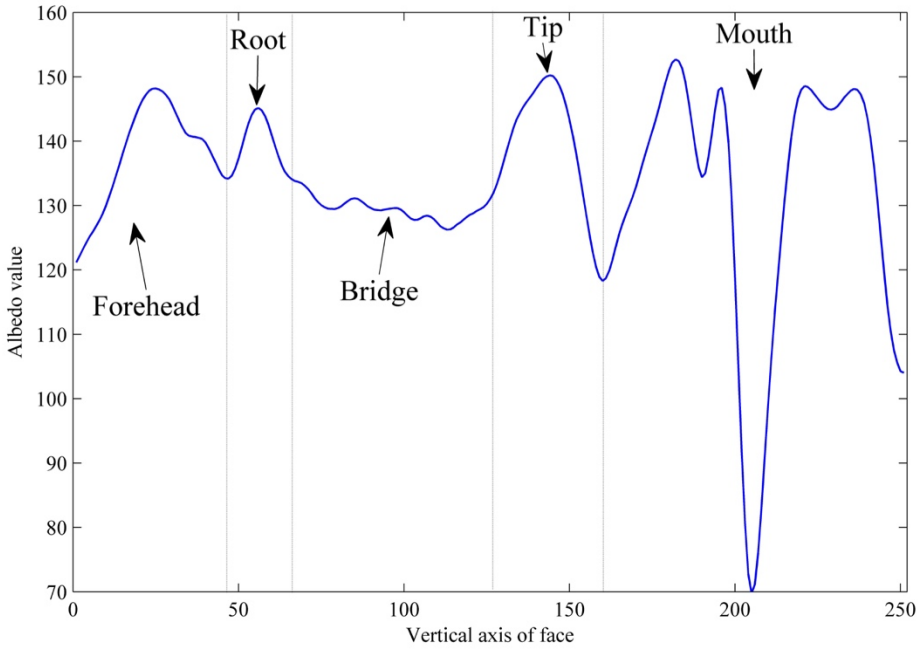


Figure 7.18: The central profile of the human face on the Photoface albedo map. The extrema points marked on the curve are the forehead, nasal root, nasal tip and month.

7.4.2 Alar Grooves and Subnasal Detection

With a reference of the nose tip, the alar grooves and subnasal can be localized by using the method proposed in Section 7.2. This is less complicated for the captures reconstructed from photometric stereo as the SNs possess relatively high resolution, which provide enough effective information for the nasal landmarking. The candidate alar grooves region is found by thresholding SNs but setting different thresholds in comparison with those used in the Bosphorus and FRGC databases. It is not necessary to adjust the location of the nose tip in this step as less noise data affects the performance of tip detection in the Photoface database.

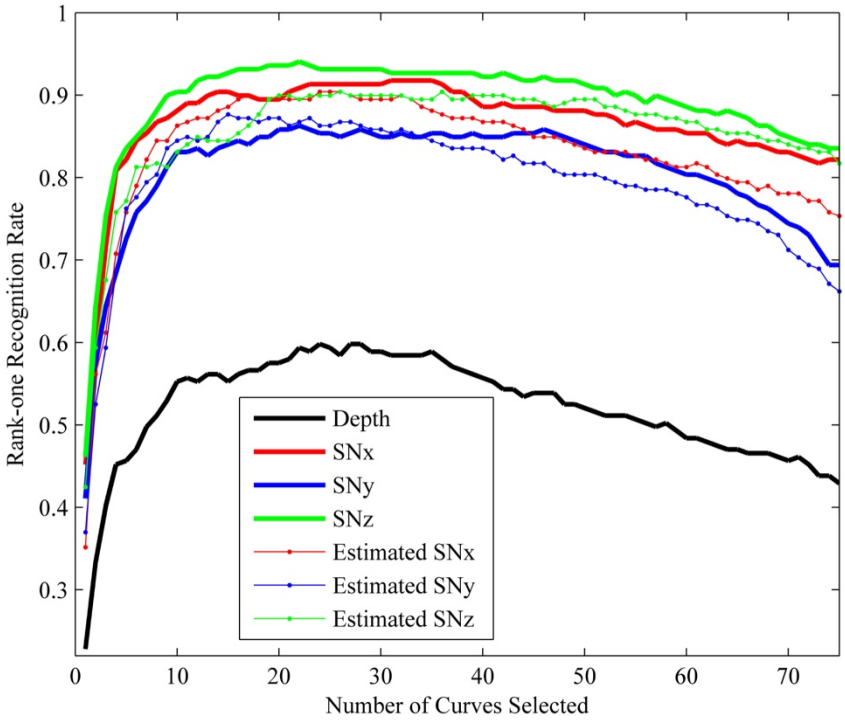
7.4.3 Recognition Performance Evaluation

In a similar method to the recognition performance evaluations on the Bosphorus and FRGC databases, using the locations of the nasal root, tip and alar grooves, the NCM algorithm proposed in [7] is applied to extract 75 nasal curves with 15 points per curve, as shown in Figure 6.1. The recognition performance evaluations of the Photoface captures employ the 3DE-VISIR database as it contains the captures obtained by both visible (VIS) light and near

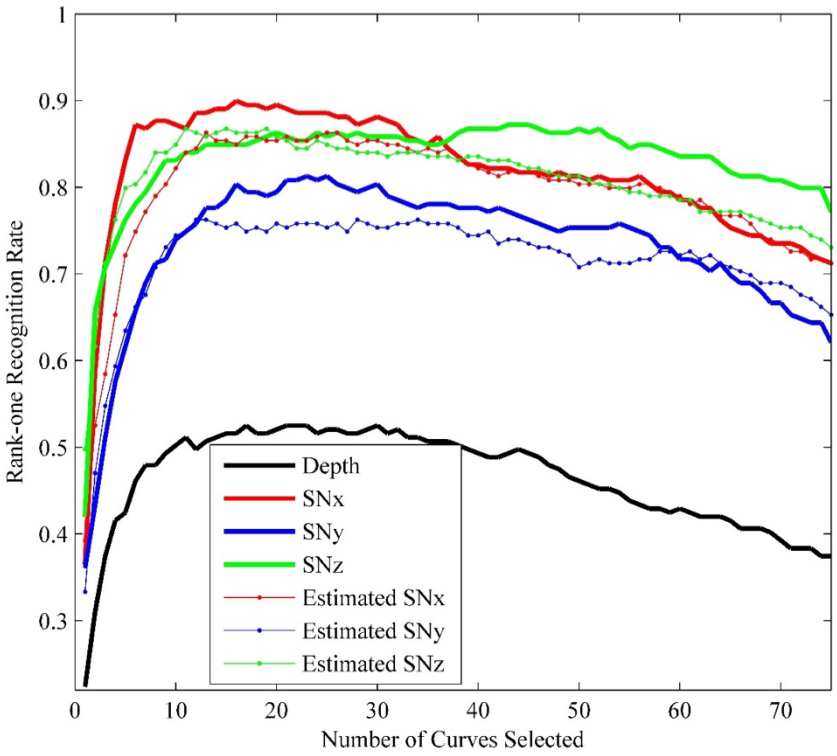
infrared (IR) light acquisition systems. For each part (VIS and IR), there are 73 subjects and three captures per subject, with the positive, neutral and negative expressions, respectively.

The recognition performance of the extracted features is first evaluated by using the LOO cross validation. To achieve the best performance, the FSFS is used for curves selection and the R_1RRs of each curve or curves combinations are calculated by LOO with city-block distance. The R_1RRs obtained from different components versus the size of the feature set is shown in Figure 7.19. The curves extracted from the surface normals achieve better recognition performances than the depth map for both the IR and VIS captures.

In addition, to explain the advantages of using the original surface normals for landmarking and feature extraction, the same feature extraction method is applied to the surface normals estimated by re-differentiating the reconstructed depth map. Generally, the R_1RRs obtained from both parts experience the decrease when the estimated surface normals are used, which mainly results from the loss of discriminative features during the depth reconstruction and surface normals estimation. Therefore, using the original surface normals has more potential to achieve good recognition performance with the low complexity feature detection and extraction.



(a) IR



(b) VIS

Figure 7.19: Rank one recognition rate against the number of nasal curves selected by the FSFS algorithm using various components. The results of surface normals and the depth reconstructed from surface normals are shown in bold. R_1RRs using the surface normals found by re-differentiating the reconstructed depth map are shown with dots.

Table 7.5: R_1RRs of the best combination of curves for SN and depth components using IR and VIS captures under 'neutral vs. non-neutral' one training sample scenario.

	R_1RR (IR)		**R_1RR (VIS)**	
	Positive	**Negative**	**Positive**	**Negative**
Depth	49.31%	42.47%	42.47%	47.95%
SN_x	80.82%	84.93%	84.93%	87.67%
SN_y	84.93%	72.60%	82.19%	68.49%
SN_z	82.19%	83.56%	80.82%	73.97%

Instead of using the LOO cross validation, 'neutral vs. non-neutral' one training sample scenario is also investigated in both parts. The captures with positive or negative expression are compared with their neutral capture, respectively. As the matching results shown in Table 7.5, the surface normals components outperform the depth for each part and kind of expressions. Also, the NCM feature extraction by using the proposed landmarking algorithm produces

similar recognition performance for both IR and VIS parts only with varying results for different expressions. Matching the negative captures from the VIS part achieves the best R_1RRs of 87.67%, which means that only several curves with 15 points per curve on the SN_x map can produce high recognition performance.

To extract more discriminative feature and achieve a better recognition performance, the newly developed spherical patches in [143] are also applied to IR part. Specifically, using the detected landmarks, tip, root and alar grooves, 21 new landmarks as shown in Figure 7.20 are found on the nasal region by dividing the horizontal lines that connect the landmarks. Considering each landmark as the centre, a sphere is intersected with the nasal region on three components of surface normals and the histograms of the inner parts are calculated.

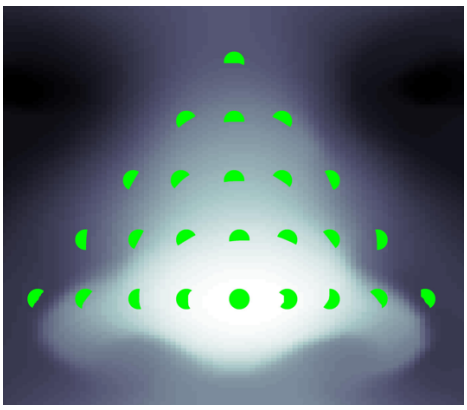


Figure 7.20: 25 landmarks used for spherical patches extraction.

In this implement, the radius is set to 11mm and there are 41 bins in each histogram. Combine all the spherical patches extracted from surface normals, there are 75 patches in total for each capture and FSFS is applied to select the most discriminative features. It achieves 97.26% of R_1RR when only 24 patches are selected for matching.

7.5 Conclusion

The main contribution of this paper is the new landmarking algorithm proposed for captures from different 3D acquisition systems, including laser scanners, structured light and photometric stereo. In addition to using the other descriptors calculated from the depth information, this landmarking algorithm employs the surface normals to quickly find the candidate regions and further localizes the main nasal landmarks by using profile signatures and thresholded surface normals maps. Finally, five landmarks, including the nasal tip, root, subnasal and alar grooves are automatically detected on the nasal region.

The proposed landmarking algorithm is evaluated on the Bosphorus and FRGC captures with higher accuracy depth map as well as the Photoface captures. The results tested on the Bosphorus and FRGC captures demonstrate that the detected landmarks possess higher within-class consistency and the features extracted from them can generate higher recognition performance using both LOO cross validation and KFA classifier in comparison with the method proposed in [7].

For the Photoface captures, there is no literature utilizing the landmarking on this kind of captures, even though they have much potential to generate good recognition performance and contribute the real world applications. The proposed nasal landmarking algorithm only utilizes the original surface normals and albedo maps of the Photoface captures and fills this gap in the landmarking literature. Both the results of the LOO cross validation and neutral vs. non-neutral matching can show the effectiveness of the proposed landmarking algorithm.

Chapter 8 Conclusions

8.1 Summary

This thesis has provided an investigation into the benefits of using the nasal and its environs for expression invariant 3D face recognition as well as an exploration of various types of captures from 3D acquisition systems. The background knowledge, basic techniques and the main challenges are first explained in Chapter 1 to give the motivation of this thesis:

- Expression robust discriminative feature extraction on the nasal and adjoining cheek regions. As one of the main challenging issues in face recognition, expression variations might adversely affect the within-class similarity of extracted features. To address this problem, this thesis aims at selecting some relatively stable structures on the 3D captures to extract the expression robust discriminative features, such as the nasal region and its environs.
- Investigating the use of reconstructed captures from photometric stereo. Another main motivation of this topic is improving the real world application possibility, for example of utilising reconstructed captures from photometric stereo. Although using the Photoface captures has a lot of advantages and promising application prospects, region based expression robust feature extraction algorithms are still unexplored. It is interesting to apply the well-designed face recognition algorithms to the Photoface captures or investigate new techniques on the basis of their characteristics.
- Landmarking on various types of captures from 3D acquisition systems. For most region based algorithms, landmarking is the most significant step, whose accuracy will directly determine the effectiveness of feature detection and extraction. For captures with high accuracy depth maps, both the efficiency and effectiveness of the landmarking should be improved. For the Photoface captures, region based face recognition algorithms are still unexplored due to the lack of effective landmarking algorithms. Therefore, this thesis aims to fill in the gap in the 3D landmark detection literature and further improve the recognition performance.

In Chapter 2, an overview of 3D face recognition algorithms is first provided, including preprocessing, landmarking, features extraction, postprocessing, matching and decision marking. The recognition algorithm for addressing the expression variations and feature

extraction on the nasal and adjoining cheek regions are further explained. Three whole face evaluations on the Bosphorus database are also provided in Chapter to demonstrate the advantages of using 3D nasal and adjoining cheek regions.

A classification methodology and review of 3D face databases are introduced in Chapter 3 to provide researchers with meaningful databases on which to evaluate their algorithms. Considering the various characteristics of 3D databases, a classification methodology has been proposed to categorise them, which provides a useful tool to help choose appropriate databases to address specific problems. On the basis of this overview, FRGC, Bosphorus and Photoface databases are finally chosen in the following chapters for the propose of better evaluating the proposed algorithms.

In Chapter 4, instead of using the whole face or other complicated shape descriptors, matching nasal curves is an effective algorithm for expression robust 3D face recognition [7]. Inspired by this work, this thesis takes effort to improve the recognition performance by finding more discriminative curves or features with low complexity, which can be proposed as a pattern rejector for 3D face recognition. To be more specific, the target region is first enlarged to the nasal environs and instead of using 28 nasal curves in [7], only 4 curves are finally selected by FSFS to generate the feature space. Also, the number of points on each curve is reduced from '50' to '15', resulting in only 60 points on the target region to build the feature set. The recognition performance tested on both Bosphorus and FRGC databases is improved, even though the dimensionality has been significantly reduced, from 1400 to 60, which mainly results from the contribution of discriminative features extracted from the nasal environs.

In Chapter 5, to further investigate the discriminative features on the nasal and its environs, a novel low complexity local shape descriptor for expression robust 3D face recognition is proposed, which is proven to be more effective than matching curves on the same target region. Instead of using selected curves on the depth map, a new shape descriptor which is inspired by Local Binary Patterns is proposed that uses depth difference to form its discriminative features. This newly proposed shape descriptor is applied to the nasal and adjoining cheeks of depth and three components of surface normals. To build a stronger classifier, a comprehensive analysis of the discriminative features extraction on multi-component and multi-scale horizontal regions or small patches is provided in Chapter 5. Using this proposed descriptor to extract features on the nasal and adjoining cheek regions results in a very small feature set, which has much potential for real world applications. For classification and matching, the KFA classifier with

the cosine distance produces a R_1RR of 97.76% and an EER of 1.32% for the Bosphorus dataset, which is competitive with the algorithms that are more computationally intensive and require a larger feature set.

The main purpose is extracting discriminative features from the Photoface captures using the 3D captures with higher accuracy depth maps, in Chapters 4 and 5. In order to investigate the recognition performance of reconstructed captures from photometric stereo, Chapter 6 evaluates the performance using the well-designed nasal curves matching algorithm. Subsets of FRGC v.2, Bosphorus and Photoface databases, 360 captures of 18 subjects, are selected to enable a direct performance comparison. In addition, the 3DE-VISIR database is used for one training sample scenario evaluation. The advantage of using this database is that each subject is obtained under both visible and near infrared light with two different expressions and a neutral capture, resulting in 6 captures for each identity. Recognition performance evaluations on this kind of captures contributes to real world face capture as the Photoface device is more effective and efficient and does not require candidates' cooperation.

In Chapter 7, a new landmarking algorithm is proposed to improve the nasal landmarks' accuracy and hence the robustness of extracted features. This new landmarking strategy is originally inspired by the successful applications of using the surface normals for feature extraction, which can provide an additional and effective representation for 3D surface geometry structure analysis. Therefore, in Chapter 7, the surface normals are further investigated for the roll and yaw rotations calibration and nasal landmarks localization. Using the thresholded surface normal maps and facial profile signatures, five landmarks (the nose tip, root, two alar grooves and subnasal) are robustly detected on the well-aligned 3D nasal region.

This new landmarking strategy provides an effective way to localize the key landmarks on human nose and can be applied to depth maps from captures obtained using laser scanner and photometric stereo. The proposed landmarking algorithm is tested on the Bosphorus and FRGC databases which contain a large number of expression variations. The results show that the detected landmarks demonstrate high within-class consistency and good recognition performance under expressions. For the Photoface captures, the main nasal landmarks are robustly detected, which is the first landmarking work on this kind of 3D captures. Instead of using the reconstructed depth information that suffers more distortion, the proposed landmarking approach utilises the surface normals, to achieve more accurate landmarking and high recognition performances.

In conclusion, this thesis is dedicated to extracting relatively expression robust features on the nasal and its environs for 3D face recognition using various 3D representations of captures from different 3D acquisition systems, which consists of the following innovative steps.

- An expression robustness analysis using both depth and surface normals information is explained for the first time.
- An overview and a classification methodology of the public 3D face database are first introduced to provide an appropriate reference for the choice of 3D face databases.
- A low complexity pattern rejector for expression robust 3D face recognition
- A novel multi-scale and multi-component local shape descriptor for expression robust discriminative features extraction.
- A comparison of the feature extraction and recognition performances using the Bosphorus, FRGC and Photoface databases.
- A novel 3D facial landmarking approach for the Photoface captures.
- Facial roll and yaw rotations calibration by detected nasal bridge and landmarks.
- Recognition performances evaluation on the 3DE-VISIR database.

8.2 Discussion and Future Work

The algorithms proposed in this thesis could be improved in a number of different aspects. Some examples are given in the following subsections.

8.2.1 The Improvement of Denoising

It is very challenging to find an appropriate denoising algorithms as the remaining noise might significantly affect the accuracy of post-processing and over-smoothed captures might lose their discriminative features. It is hard to propose a generalized and effective approach to compromise these two aspects. Median filtering has been shown to be the best choice, which has been tested by several holistic algorithms [135]. However, for region based face recognition algorithms, it is still not clear which filters are more effective.

The original output of most 3D acquisition systems is the point clouds from which other 3D representations, such as surface normals and curvatures, can be estimated. Therefore, any noise

not removed from the point clouds might be amplified in the other representations. In Chapter 6, the median and Gabor filters are applied to further smooth the normals and shape index maps, which have significantly improved the performance of single nasal curve matching. However, how to extract more discriminative features on the smoothed maps or employ more effective filters is still very challenging and worth to further investigate.

8.2.2 Feature Extraction on Surface Normals

In Chapter 7, the recognition results show that the original SNs outperform the SNs estimated by the reconstructed depth information, which demonstrates the potential of the proposed landmarking approach and also the use of SNs for feature space manipulation. However, the recognition algorithms used in performances evaluations are originally proposed for the depth map, which might not the most appropriate approach for the Photoface captures. Using the newly detected landmarks on the nasal region, the new feature extraction algorithms are required to apply to the SNs maps. In Chapter 6, different variances or frequencies of Gabor filters have been used to denoise the SNs, which provide a great contribution to the improvement of recognition performances. Therefore, in future work, the Gabor wavelets will be applied to the feature extraction on the Photoface captures as they can smooth the original surface normals maps and various discriminative features can be extracted from their different scales or frequencies.

8.2.3 More Effective Feature Selection

Feature selection is very crucial in feature space manipulation, which determines the dimensionality and discrimination of the feature set for recognition. Therefore, this step might directly affect the face recognition performance. In this thesis, the FSFS method is primarily used in feature selection and the demonstration of the recognition rate versus the number of curves in the leave-one-out cross validation. However, the FSFS does not evaluate all the curves combinations, which cannot produce the global optimal features set. Although it is clear to know the selected curves in each iteration by FSFS, it is still interesting to employ some other feature selection methods, for example of GAs, to further enhance the recognition performance.

8.2.4 One Training Sample Scenario Evaluation

Most classification evaluations in this thesis use at least one capture for each subject in the training set, which has higher probability to produce good recognition performance as the

features for matching are learned from more captures. Another more challenging scenario is when each person only has one capture in the gallery, which requires the matching features should possess high discrimination. In Chapter 6, this 'One Training Sample' scenario has been evaluated on the 3DE-VISIR database by matching the nasal curves. However, most features extracted in this thesis have not been evaluated under this scenario due to the lack of effective feature selection method. Therefore, one area of future work will focus on building an effective subset from the original features using some feature selection methods or fusing other features to build a new feature vector with more discriminative information for matching.

8.2.5 Artificially Creating Facial Captures with Expressions

Instead of extracting more discriminative features for one-to-one matching, the recognition performance can be improved by adding more artificial training captures of each subject. To artificially create new captures, all kinds of facial variations caused by various expressions and occlusions can be first learned from the predefined facial captures. For example, various expressions and occlusions have been labelled in the Bosphorus database and the variations can be learned from it. These learned facial variations are then added to the single neutral capture of each subject in the training set, resulting in many captures with different kinds of variations. With the increasing numbers of training captures, the recognition performance can be significantly improved. A related work has been done in my PhD study and the recognition performance only obtained incremental improvement as it is very challenging to model different kinds of variations. More effective methods for artificially creating the facial captures are encouraged in the future work.

List of publications arising from this thesis

1. J. Gao, M. Emambakhsh, and A. Evans, "*A Low Dimensionality Expression Robust Rejector for 3D Face Recognition*", in Proceedings of IEEE International Conference on Pattern Recognition (ICPR 2014).

2. J. Gao, and A. Evans, "*Expression robust 3D face recognition by matching multi-component local shape descriptors on the nasal and adjoining cheek regions*", in Proceedings of 11th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2015).

3. J. Gao, and A. Evans, "*Using 3D Representations of the Nasal Region for Improved Landmarking and Expression Robust Recognition*", in Proceedings of 7th UK Computer Vision Student Workshop (BMVW 2015).

4. M. Emambakhsh, J. Gao, and A. Evans, "*An evaluation of denoising algorithms for 3D face recognition*," in Proceedings of IET 5th International Conference on Imaging for Crime Detection and Prevention (ICDP 2013).

5. M. Emambakhsh, J. Gao, and A. Evans, "*Noise modelling for denoising and 3D face recognition algorithms performance evaluation*", IET Computer Vision, 9 (5), pp. 741-749.

References

[1] K. B. Raja, R. Raghavendra, V. K. Vemuri, and C. Busch, "Smartphone based visible iris recognition using deep sparse filtering," *Pattern Recognition Letters,* vol. 57, pp. 33-42, 2015.

[2] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: a survey," *Proceedings of the IEEE,* vol. 83, pp. 705-741, 1995.

[3] A. Moorhouse, A. Evans, G. A. Atkinson, J. Sun, and M. L. Smith, "The nose on your face may not be so plain: Using the nose as a biometric," in *3rd International Conference on Imaging for Crime Detection and Prevention*. vol. 2009, ed: Institution of Engineering and Technology, 2009.

[4] K. I. Chang, W. Bowyer, and P. J. Flynn, "Multiple Nose Region Matching for 3D Face Recognition under Varying Facial Expression," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28, pp. 1695-1700, 2006.

[5] A. S. Mian, M. Bennamoun, and R. Owens, "An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 29, pp. 1927-1943, 2007.

[6] H. Li, D. Huang, J.-M. Morvan, L. Chen, and Y. Wang, "Expression-robust 3D face recognition via weighted sparse representation of multi-scale and multi-component local normal patterns," *Neurocomputing,* vol. 133, pp. 179-193, 2014.

[7] M. Emambakhsh, A. Evans, and M. Smith, "Using nasal curves matching for expression robust 3D nose recognition," presented at the IEEE International Conference on Biometrics: Theory, Applications and Systems 2013.

[8] W. Yueming, T. Xiaoou, L. Jianzhuang, P. Gang, and X. Rong, "3D Face Recognition by Local Shape Difference Boosting," presented at the European Conference on Computer Vision, 603-616, 2008.

[9] S. Zafeiriou, G. A. Atkinson, M. F. Hansen, W. A. P. Smith, V. Argyriou, M. Petrou*, et al.*, "Face Recognition and Verification using Photometric Stereo: The Photoface Database and a Comprehensive Evaluation," *IEEE Transactions on Information Forensics and Security,* vol. 8, pp. 121-135, 2012.

[10] M. F. Hansen, G. A. Atkinson, L. N. Smith, and M. L. Smith, "3D face reconstructions from photometric stereo using near infrared and visible light," *Computer Vision and Image Understanding,* vol. 114, pp. 942-951, 2010.

[11] R. T. Frankot and R. Chellappa, "A method for enforcing integrability in shape from shading algorithms," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 10, pp. 439-451, 1988.

[12] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, C. Jin, K. Hoffman*, et al.*, "Overview of the face recognition grand challenge," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, pp. 947-954

[13] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur*, et al.*, "Bosphorus Database for 3D Face Analysis," in *Biometrics and Identity Management*, ed: Springer-Verlag, 2008, pp. 47-56.

[14] T. Huynh, R. Min, and J.-L. Dugelay, "An efficient LBP-based descriptor for facial depth images applied to gender recognition using RGB-D face data," in *Asian Conference on Computer Vision*, 2012, pp. 133-145.

[15] J. Shi, A. Samal, and D. Marx, "How effective are landmarks and their geometry for face recognition?," *Computer Vision and Image Understanding,* vol. 102, pp. 117-133, 2006.

[16] C. Creusot, N. Pears, and J. Austin, "A Machine-Learning Approach to Keypoint Detection and Landmarking on 3D Meshes," *International Journal of Computer Vision,* vol. 102, pp. 146-179, 2013.

[17] S. Zafeiriou, M. Hansen, G. Atkinson, V. Argyriou, M. Petrou, M. Smith*, et al.*, "The Photoface database," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2011, pp. 132-139.

[18] M. Hansen, "3D face recognition using photometric stereo," PhD, Faculty of the Environment and Technology, University of the West of England, 2012.

[19] B. Gökberk, M. O. İrfanoğlu, and L. Akarun, "3D shape-based face representation and feature extraction for face recognition," *Image and Vision Computing,* vol. 24, pp. 857-869, 2006.

[20] H. Mohammadzade and D. Hatzinakos, "Iterative Closest Normal Point for 3D Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 35, pp. 381-397, 2013.

[21] *Biometric Consortium*. Available: http://www.biometrics.org/

[22] A. K. Jain, A. Ross, and S. Pankanti, "Biometrics: a tool for information security," *Information Forensics and Security, IEEE Transactions on,* vol. 1, pp. 125-143, 2006.

[23] A. Jain, L. Hong, and S. Pankanti, "Biometric identification," *Communications of the ACM,* vol. 43, pp. 90-98, 2000.

[24] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 14, pp. 4-20, 2004.

[25] A. K. Jain, A. Ross, and S. Pankanti, "Biometrics: a tool for information security," *IEEE Transactions on Information Forensics and Security,* vol. 1, pp. 125-143, 2006.

[26] S. Z. Li and A. K. Jain, *Handbook of face recognition*: springer, 2011.

[27] K. W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *Computer Vision and Image Understanding,* vol. 101, pp. 1-15, 2006.

[28] A. Bronstein, M. Bronstein, and R. Kimmel, "Expression-invariant 3D face recognition," in *Audio-and Video-Based Biometric Person Authentication*, 2003, pp. 62-70.

[29] C. Hesher, A. Srivastava, and G. Erlebacher, "A novel technique for face recognition using range imaging," in *Seventh International Symposium on Signal Processing and Its Applications*, 2003, pp. 201-204.

[30] A. Scheenstra, A. Ruifrok, and R. Veltkamp, "A survey of 3D face recognition methods," in *Audio-and Video-Based Biometric Person Authentication*, 2005, pp. 325-345.

[31] G. Medioni and R. Waupotitsch, "Face modeling and recognition in 3-D," in *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, 2003, pp. 232-233.

[32] G. G. Gordon, "Face recognition based on depth and curvature features," in *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on*, 1992, pp. 808-810.

[33] N. Alyuz, B. Gokberk, H. Dibeklioglu, and L. Akarun, "Component-based registration with curvature descriptors for expression insensitive 3d face recognition," in *8th IEEE International Conference on Automatic Face & Gesture Recognition*, 2008, pp. 1-6.

[34] J. Gao, M. Emambakhsh, and A. N. Evans, "A Low Dimensionality Expression Robust Rejector for 3D Face Recognition," in *22nd International Conference on Pattern Recognition (ICPR)*, 2014, pp. 506-511.

[35] J. J. Koenderink and A. J. van Doorn, "Surface shape and curvature scales," *Image and Vision Computing,* vol. 10, pp. 557-564, 1992.

[36] Z. Zhu, J. Gao, and H. Yu, "Face Detection Based on Fractal and Complexion Model in the Complex Background," in *International Workshop on Chaos-Fractals Theories and Applications (IWCFTA)*, 2010, pp. 491-495.

[37] A. Colombo, C. Cusano, and R. Schettini, "3D face detection using curvature analysis," *Pattern Recognition,* vol. 39, pp. 444-455, 2006.

[38] P. Szeptycki, M. Ardabilian, and C. Liming, "A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking," in *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, 2009, pp. 1-6.

[39] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis*, et al.*, "Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 29, pp. 640-649, 2007.

[40] P. Perakis, G. Passalis, T. Theoharis, and I. Kakadiaris, "3D Facial Landmark Detection under Large Yaw and Expression Variations," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 35, pp. 1552 - 1564, 2013.

[41] G. Passalis, P. Perakis, T. Theoharis, and I. A. Kakadiaris, "Using Facial Symmetry to Handle Pose Variations in Real-World 3D Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 33, pp. 1938-1951, 2011.

[42] B. Efraty, E. Bilgazyev, S. Shah, and I. A. Kakadiaris, "Profile-based 3D-aided face recognition," *Pattern Recognition,* vol. 45, pp. 43-53, 2012.

[43] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "An Evaluation of Multimodal 2D+3D Face Biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, pp. 619-624, 2005.

[44] Y. Lei, M. Bennamoun, and A. A. El-Sallam, "An efficient 3D face recognition approach based on the fusion of novel local low-level features," *Pattern Recognition,* vol. 46, pp. 24-37, 2013.

[45] I. C. Kyong, K. W. Bowyer, and P. J. Flynn, "Adaptive Rigid Multi-region Selection for Handling Expression Variation in 3D Face Recognition," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 2005, pp. 157-157.

[46] W. Yueming, L. Jianzhuang, and T. Xiaoou, "Robust 3D Face Recognition by Local Shape Difference Boosting," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 32, pp. 1858-1870, 2010.

[47] C. Xu, S. Li, T. Tan, and L. Quan, "Automatic 3D face recognition from depth and intensity Gabor features," *Pattern Recognition,* vol. 42, pp. 1895-1905, 2009.

[48] R. Sala Llonch, E. Kokiopoulou, I. Tošić, and P. Frossard, "3D face recognition with sparse spherical representations," *Pattern Recognition,* vol. 43, pp. 824-834, 2010.

[49] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama, "3D Face Recognition Under Expressions, Occlusions and Pose Variations," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 35, pp. 2270 - 2283, 2013.

[50] J. Wang, X. Zhang, and Z. Yu, "A cascaded approach for feature-preserving surface mesh denoising," *Computer-Aided Design,* vol. 44, pp. 597-610, 2012.

[51] H. Tang, B. Yin, Y. Sun, and Y. Hu, "3D face recognition using local binary patterns," *Signal Processing,* vol. 93, pp. 2190-2198, 2013.

[52] X. Li and F. Da, "Efficient 3D face recognition handling facial expression and hair occlusion," *Image and Vision Computing,* vol. 30, pp. 668-679, 2012.

[53] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn, "Using multi-instance enrollment to improve performance of 3D face recognition," *Computer Vision and Image Understanding,* vol. 112, pp. 114-125, 2008.

[54] Y. Lei, M. Bennamoun, M. Hayat, and Y. Guo, "An Efficient 3D Face Recognition Approach using Local Geometrical Signatures," *Pattern Recognition,* vol. 47, 2013.

[55] Y. Ming and Q. Ruan, "Robust sparse bounding sphere for 3D face recognition," *Image and Vision Computing,* vol. 30, pp. 524-534, 2012.

[56] C. C. Queirolo, L. Silva, O. R. P. Bellon, and M. P. Segundo, "3D Face Recognition Using Simulated Annealing and the Surface Interpenetration Measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 32, pp. 206-219, 2010.

[57] E. Murphy-Chutorian and M. M. Trivedi, "Head Pose Estimation in Computer Vision: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 31, pp. 607-626, 2009.

[58] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 14, pp. 239-256, 1992.

[59] A. U. Batur and M. H. Hayes, "Adaptive active appearance models," *IEEE Transactions on Image Processing,* vol. 14, pp. 1707-1721, 2005.

[60] M. Pamplona Segundo, L. Silva, O. R. P. Bellon, and C. C. Queirolo, "Automatic Face Segmentation and Facial Landmark Detection in Range Images," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,* vol. 40, pp. 1319-1330, 2010.

[61] J. Gao and A. N. Evans, "Expression robust 3D face recognition by matching multi-component local shape descriptors on the nasal and adjoining cheek regions," presented at the 11th IEEE International Conference on Automatic Face and Gesture Recogntion, Ljublijana, Slovenia, 2015.

[62] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*: Addison- Wesley 1992.

[63] D. Colbry, G. Stockman, and A. Jain, "Detection of Anchor Points for 3D Face Veri.cation," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 2005, pp. 118-118.

[64] J. D'Hose, J. Colineau, C. Bichon, and B. Dorizzi, "Precise Localization of Landmarks on 3D Faces using Gabor Wavelets," in *IEEE International Conference on Biometrics: Theory, Applications, and Systems*, 2007, pp. 1-6.

[65] N. Pears, T. Heseltine, and M. Romero, "From 3D Point Clouds to Pose-Normalised Depth Maps," *International Journal of Computer Vision,* vol. 89, pp. 152-176, 2010.

[66] M. P. Segundo, C. Queirolo, O. R. P. Bellon, and L. Silva, "Automatic 3D facial segmentation and landmark detection," in *14th International Conference on Image Analysis and Processing*, 2007, pp. 431-436.

[67] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn, "Rotated Profile Signatures for robust 3D feature detection," in *8th IEEE International Conference on Automatic Face & Gesture Recognition*, 2008, pp. 1-7.

[68] A. Mian, M. Bennamoun, and R. Owens, "Automatic 3d face detection, normalization and recognition," in *Third International Symposium on 3D Data Processing, Visualization, and Transmission*, 2006, pp. 735-742.

[69] S. Berretti, A. D. Bimbo, and P. Pala, "Recognition of 3D faces with missing parts based on profile networks," presented at the ACM Workshop on 3D Object Retrieval, Firenze, Italy, 2010.

[70] Z. Xi, E. Dellandrea, C. Liming, and I. A. Kakadiaris, "Accurate Landmarking of Three-Dimensional Facial Data in the Presence of Facial Expressions and Occlusions Using a Three-Dimensional Statistical Facial Feature Model," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,* vol. 41, pp. 1417-1428, 2011.

[71] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *Acm Computing Surveys (CSUR),* vol. 35, pp. 399-458, 2003.

[72] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience,* vol. 3, pp. 71-86, 1991.

[73] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 19, pp. 711-720, 1997.

[74] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,,* vol. 31, pp. 210-227, 2009.

[75] T. Zhang, Y. Y. Tang, B. Fang, Z. Shang, and X. Liu, "Face recognition under varying illumination using gradientfaces," *IEEE Transactions on Image Processing,* vol. 18, pp. 2599-2606, 2009.

[76] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, pp. 328-340, 2005.

[77] M. S. Bartlett, M. H. Lades, and T. J. Sejnowski, "Independent component representations for face recognition," in *Proceedings of SPIE*, 1998, p. 528.

[78] P. J. Phillips, *Support vector machines applied to face recognition*: Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems II, 1998.

[79] C. Liu and H. Wechsler, "Evolutionary pursuit and its application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, pp. 570-582, 2000.

[80] L. Ballihi, B. Ben Amor, M. Daoudi, A. Srivastava, and D. Aboutajdine, "Boosting 3-D-Geometric Features for Efficient Face Recognition and Gender Classification," *IEEE Transactions on Information Forensics and Security,* vol. 7, pp. 1766-1779, 2012.

[81] G. Zhang, X. Huang, S. Z. Li, Y. Wang, and X. Wu, "Boosting local binary pattern (LBP)-based face recognition," in *Advances in Biometric Person Authentication*, ed: Springer, 2005, pp. 179-186.

[82] C. I. Podilchuk and X. Zhang, "Face recognition using DCT-based feature vectors," ed: Google Patents, 1998.

[83] V. Sˇtruc and N. Pavesˇic, "The complete gabor-fisher classifier for robust face recognition," *EURASIP J. Adv. Signal Process,* vol. 2010, pp. 1-13, 2010.

[84] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems*, 2001, pp. 556-562.

[85] N. Alyuz, B. Gokberk, and L. Akarun, "Regional Registration for Expression Resistant 3-D Face Recognition," *IEEE Transactions on Information Forensics and Security,,* vol. 5, pp. 425-440, 2010.

[86] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin, "Static and dynamic 3D facial expression recognition: A comprehensive survey," *Image and Vision Computing,* vol. 30, pp. 683-697, 2012.

[87] P. Lemaire, B. Ben Amor, M. Ardabilian, L. Chen, and M. Daoudi, "Fully automatic 3D facial expression recognition using a region-based approach," in *Joint ACM workshop on Human gesture and behavior understanding*, 2011, pp. 53-58.

[88] A. Maalej, B. B. Amor, M. Daoudi, A. Srivastava, and S. Berretti, "Shape analysis of local facial patches for 3D facial expression recognition," *Pattern Recognition,* vol. 44, pp. 1581-1589, 2011.

[89] K. S. Park and N. S. Lee, "A three-dimensional Fourier descriptor for human body representation/reconstruction from serial cross sections," *Computers and Biomedical Research,* vol. 20, pp. 125-140, 1987.

[90] B. K. Horn, "Extended gaussian images," *Proceedings of the IEEE,* vol. 72, pp. 1671-1686, 1984.

[91] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 21, pp. 433-449, 1999.

[92] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters,* vol. 15, pp. 1119-1125, 1994.

[93] I. A. Gheyas and L. S. Smith, "Feature subset selection in large dimensionality domains," *Pattern Recognition,* vol. 43, pp. 5-13, 2010.

[94] Z. Sun, G. Bebis, and R. Miller, "Object detection using feature subset selection," *Pattern Recognition,* vol. 37, pp. 2165-2176, 2004.

[95] C. M. Bishop, *Neural networks for pattern recognition*: Oxford university press, 1995.

[96] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*: John Wiley & Sons, 2012.

[97] C. Xu, T. Tan, S. Li, Y. Wang, and C. Zhong, "Learning effective intrinsic features to boost 3d-based face recognition," presented at the Proceedings of the 9th European conference on Computer Vision - Volume Part II, Graz, Austria, 2006.

[98] J. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, and S. Z. Li, "Ensemble-based discriminant learning with boosting for face recognition," *IEEE Transactions on Neural Networks,* vol. 17, pp. 166-178, 2006.

[99] B. Achermann, X. Jiang, and H. Bunke, "Face recognition using range images," in *International Conference on Virtual Systems and MultiMedia*, 1997, pp. 129-136.

[100] F. Tsalakanidou, S. Malassiotis, and M. G. Strintzis, "Integration of 2D and 3D images for enhanced face authentication," in *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 266-271.

[101] X. Lu, D. Colbry, and A. K. Jain, "Matching 2.5 D scans for face recognition," in *Biometric Authentication*, ed: Springer, 2004, pp. 30-36.

[102] Y. Lee, H. Song, U. Yang, H. Shin, and K. Sohn, "Local Feature Based 3D Face Recognition," in *Audio- and Video-Based Biometric Person Authentication*. vol. 3546, ed: Springer Berlin Heidelberg, 2005, pp. 909-918.

[103] D. P. Huttenlocher, G. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 15, pp. 850-863, 1993.

[104] T. D. Russ, M. W. Koch, and C. Q. Little, "A 2D Range Hausdorff Approach for 3D Face Recognition," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 2005, p. 169.

[105] W. Deng, J. Hu, and J. Guo, "Extended SRC: Undersampled face recognition via intraclass variant dictionary," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 34, pp. 1864-1870, 2012.

[106] Y. Bengio, "Learning deep architectures for AI," *Foundations and trends® in Machine Learning,* vol. 2, pp. 1-127, 2009.

[107] D. Zhang and W. Zuo, "Computational intelligence-based biometric technologies," *IEEE Computational Intelligence Magazine,* vol. 2, pp. 26-36, 2007.

[108] S.-H. Lin, S.-Y. Kung, and L.-J. Lin, "Face recognition/detection by probabilistic decision-based neural network," *IEEE Transactions on Neural Networks,* vol. 8, pp. 114-132, 1997.

[109] M. J. Er, S. Wu, J. Lu, and H. L. Toh, "Face recognition with radial basis function (RBF) neural networks," *IEEE Transactions on Neural Networks,* vol. 13, pp. 697-710, 2002.

[110] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE Transactions on Neural Networks,* vol. 8, pp. 98-113, 1997.

[111] D. Smeets, P. Claes, J. Hermans, D. Vandermeulen, and P. Suetens, "A Comparative Study of 3-D Face Recognition Under Expression Variations," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews,* vol. 42, pp. 710-727, 2012.

[112] M. Emambakhsh and A. Evans, "Self-dependent 3D face rotational alignment using the nose region," presented at the 4th International Conference on Imaging for Crime Detection and Prevention, London, 2011.

[113] H. Dibeklioğlu, B. Gökberk, and L. Akarun, "Nasal region-based 3D face recognition under pose and expression variations," in *Advances in Biometrics*, ed: Springer, 2009, pp. 309-318.

[114] H. Drira, B. B. Amor, M. Daoudi, and A. Srivastava, "Nasal region contribution in 3D face biometrics using shape analysis framework," in *Advances in Biometrics*, ed: Springer, 2009, pp. 357-366.

[115] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, pp. 971-987, 2002.

[116] D. Huang, G. Zhang, M. Ardabilian, Y. Wang, and L. Chen, "3D face recognition using distinctiveness enhanced facial representations and local feature hybrid matching," in *IEEE International Conference on Biometrics: Theory Applications and Systems* 2010, pp. 1-7.

[117] D. Huang, M. Ardabilian, Y. Wang, and L. Chen, "3-D face recognition using eLBP-based facial description and local feature hybrid matching," *IEEE Transactions on Information Forensics and Security,* vol. 7, pp. 1551-1565, 2012.

[118] X. Lu and A. K. Jain, "Integrating range and texture information for 3D face recognition," in *Seventh IEEE Workshops on Application of Computer Vision*, 2005, pp. 156-163.

[119] C. Beumier and M. Acheroy, "Face verification from 3D and grey level clues," *Pattern Recognition Letters,* vol. 22, pp. 1321-1329, 2001.

[120] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn, "Using a Multi-Instance Enrollment Representation to Improve 3D Face Recognition," in *IEEE International Conference on Biometrics: Theory, Applications and Systems* 2007, pp. 1-6.

[121] A. B. Moreno and A. Sánchez, "GavabDB: a 3D Face Database," in *2nd COST Workshop on Biometrics on the Internet: Fundamentals, Advances and Applications*, 2004, pp. 77-82.

[122] C. Conde, A. Serrano, and E. Cabello, "Multimodal 2D, 2.5D & 3D Face Verification," in *IEEE International Conference on Image Processing*, 2006, pp. 2061-2064.

[123] M. a. I. S. T. B. Key and Laboratory, "The BJUT-3D large-scale chinese face database," *Beijing Univ. Technology, Beijing, China, Tech. Rep. MISKL-TR-05-FMFR-001,* 2005.

[124] T. Heseltine, N. Pears, and J. Austin, "Three-dimensional face recognition using combinations of surface feature map subspace components," *Image and Vision Computing,* vol. 26, pp. 382-396, 2008.

[125] Y. Lijun, W. Xiaozhou, S. Yi, W. Jun, and M. J. Rosato, "A 3D facial expression database for facial behavior research," in *7th International Conference on Automatic Face and Gesture Recognition*, 2006, pp. 211-216.

[126] S. Gupta, K. R. Castleman, M. K. Markey, and A. C. Bovik, "Texas 3D Face Recognition Database," in *IEEE Southwest Symposium on Image Analysis & Interpretation (SSIAI)*, 2010, pp. 97-100.

[127] A. Colombo, C. Cusano, and R. Schettini, "UMB-DB: A database of partially occluded 3D faces," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, 2011, pp. 2113-2119.

[128] C. Boehnen and P. Flynn, "Accuracy of 3D scanning technologies in a face scanning scenario," in *Fifth International Conference on 3-D Digital Imaging and Modeling*, 2005, pp. 310-317.

[129] K. Chang, K. Bowyer, and P. Flynn, "Face recognition using 2D and 3D facial data," in *ACM Workshop on Multimodal User Authentication*, 2003, pp. 25-32.

[130] T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West*, et al.*, "Performance of Geometrix ActiveID^ TM 3D Face Recognition Engine on the FRGC Data," in *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, 2005, pp. 154-154.

[131] F. Tsalakanidou, D. Tzovaras, and M. Strintzis, "Use of depth and colour eigenfaces for face recognition," *Pattern Recognition Letters,* vol. 24, pp. 1427-1435, 2003.

[132] V. Vijayan, K. W. Bowyer, P. J. Flynn, D. Huang, L. Chen, M. Hansen*, et al.*, "Twins 3D face recognition challenge," in *International Joint Conference on Biometrics (IJCB)*, 2011, pp. 1-7.

[133] P. Nair and A. Cavallaro, "3-D Face Detection, Landmark Localization, and Registration Using a Point Distribution Model," *IEEE Transactions on Multimedia,* vol. 11, pp. 611-623, 2009.

[134] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Face recognition from a single image per person: A survey," *Pattern Recognition,* vol. 39, pp. 1725-1745, 2006.

[135] M. Emambakhsh, J. Gao, and A. Evans, "An Evaluation of Denoising Algorithms for 3D Face Recognition," presented at the 5th International Conference on Imaging for Crime Detection and Prevention, London.

[136] S. Baker and S. K. Nayar, "Pattern rejection," presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1996.

[137] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science,* vol. 290, pp. 2319-2323, 2000.

[138] J. Gao, M. Emambakhsh, and A. N. Evans. *A Low Dimensionality Expression Robust Rejector for 3D Face Recognition*, 22nd International Conference on Pattern Recognition (ICPR), pp. 506-511, 2014.

[139] C. Liu, "Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28, pp. 725-737, 2006.

[140] O. Ocegueda, G. Passalis, T. Theoharis, S. K. Shah, and I. A. Kakadiaris, "UR3D-C: Linear dimensionality reduction for efficient 3D face recognition," in *International Joint Conference on Biometrics (IJCB)*, 2011, pp. 1-6.

[141] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn, "A region ensemble for 3-D face recognition," *IEEE Transactions on Information Forensics and Security,* vol. 3, pp. 62-73, 2008.

[142] H. Di, Z. Guangpeng, M. Ardabilian, W. Yunhong, and C. Liming, "3D Face recognition using distinctiveness enhanced facial representations and local feature hybrid matching," in *IEEE International Conference on Biometrics: Theory, Applications and Systems* 2010, pp. 1-7.

[143] M. Emambakhsh, "Using the 3D shape of the nose for biometric authentication," Doctor of Philosophy (PhD), Electronic & Electrical Engineering, University of Bath, Bath, UK, 2014.

[144] J. Zivanov, P. Paysan, and T. Vetter, "Facial Normal Map Capture using Four Lights: an Effective and Inexpensive Method of Capturing the Fine Scale Detail of Human Faces using Four Point Lights," presented at the Fourth International Conference on Computer Graphics Theory and Applications, Lisboa, Portugal, 2009.

[145] J. Gao and A. N. Evans, "Using 3D Representations of the Nasal Region for Improved Landmarking and Expression Robust Recognition," in *British Machine Vision Student Workshop*, 2015.